



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 131674

To: Sarvamangala Devi
Location: REM 3C18
Art Unit: 1645
Friday, September 03, 2004

Case Serial Number: 09/743674

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54

Phone: 571-272-2528

beverly.shears@uspto.gov

Search Notes

Shears, Beverly

131674

From: Devi, Sarvamangala
Sent: Wednesday, September 01, 2004 10:56 AM
To: Shears, Beverly
Subject: 09/743,674

Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 2 in application 09/743,674?

Thanx.

S. DEVI, Ph.D.
AU 1645
Rems - 3C18



Date completed: 09-03-04
Searcher: Beverly e2528
Terminal time: 23
Elapsed time: _____
CPU time: _____
Total time: 23
Number of Searches: _____
Number of Databases: 1

Search Site

____ STIC
____ CM-1
____ Pre-S

Type of Search

____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors

____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other CGN

Result No.	Query No.	Score	Query			DB	ID	Description
			Match	Length				
1	1	4404	100.0	880	3	AY44638	Aay44638 N. mening	
2	2	3434	78.0	801	6	ABU37449	Abu37449 Protein e	
3	3	2391.5	54.3	503	6	ABU37647	Abu37647 Protein e	
4	4	1122.5	25.5	244	6	ABP80077	Abp80077 N. gonorr	
5	5	1122.5	25.5	244	6	ABP78566	Abp78566 N. gonorr	
6	6	1025	23.3	275	6	ABP78509	Abp78509 N. gonorr	
7	7	1025	23.3	275	6	ABP78509	Abp78509 N. gonorr	
8	8	821.5	18.7	253	6	ABP79684	Abp79684 N. gonorr	
9	9	620.5	14.1	132	6	ABP78538	Abp78538 N. gonorr	
10	10	381.5	8.7	919	6	ABP79950	Abp79950 N. gonorr	
11	11	358	8.1	1095	6	ABP38527	Abp38527 Protein e	
12	12	319.5	7.3	2768	4	ABJ26247	Abj26247 Aspergill	
13	13	315.5	7.2	1786	2	ABB68397	Abb68397 Drosophil	
14	14	315.5	7.2	1787	5	AAW24790	Aaw24790 P. falcip	
15	15	315	7.2	1000	6	AAU96699	Aau96699 plasmodi	
16	16	313	7.1	1822	2	ABJ25647	Abj25647 Aspergill	
17	17	289.5	6.6	1468	4	AAZ27745	Aaz27745 Extracell	
18	18	289	6.6	8991	6	ABB62391	Abb62391 Drosophil	
19	19	287.5	6.5	753	4	ABU08487	Abu08487 S. pneumo	
20	20	277	6.3	2748	4	ABB68110	Abb68110 Drosophil	
21	21	276.5	6.3	2468	6	ABB58843	Abb58843 Drosophil	
22	22	276.5	6.3	2468	7	ABR64281	AbR64281 Angiogene	
23	23	276.5	6.3	2468	7	ADE62723	AdE62723 Human pro	
24	24	276.5	6.3	2468	7	ADE62719	AdE62719 Human pro	
25	25	276.5	6.3	2468	7	ADE62727	AdE62727 Human pro	
						ADef62715	ADef62715 Human pro	

XX	SQ	Sequence	880 AA;
		Query Match	100.0%; Score 4404; DB 3; Length 880;
		Best Local Similarity	100.0%; Pred. No. 1.5e-254;
		Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MPAGRLPRRCBMMTKFTDCTRSNRIQPTTHRGYLLKNNROIKLIAASVAVAAASFOAHAGL	60
DB	1	MPAGRLPRRCBMMTKFTDCTRSNRIQPTTHRGYLLKNNROIKLIAASVAVAAASFOAHAGL	60
QY	61	GGLNIQSNLDEPFPGSGITVTGEEAKALLGGSGVTVSEKGLTAKVHKLGDKXAVIASVSEQA	120
DB	61	GGLNIQSNLDEPFPGSGITVTGEEAKALLGGSGVTVSEKGLTAKVHKLGDKXAVIASVSEQA	120
QY	121	VRDPVLVFRIGAGQVREYTAI LDPVGVSPTKTSALSDGKTHRTKTAFTASSOENONAKAL	180
DB	121	VRDPVLVFRIGAGQVREYTAI LDPVGVSPTKTSALSDGKTHRTKTAFTASSOENONAKAL	180
QY	181	RKTDKDSANAAVKPAYNGKTHTVRKGBTVQIAAAIIRPKHLITLQEVADALLKANPNVSA	240
DB	181	RKTDKDSANAAVKPAYNGKTHTVRKGBTVQIAAAIIRPKHLITLQEVADALLKANPNVSA	240
QY	241	HGRLRAGSVLHI PNLRIRKABQPKPQTAAPKAETAASMPSESPSKQATVKEPKVEPEAKVAA	300
DB	241	HGRLRAGSVLHI PNLRIRKABQPKPQTAAPKAETAASMPSESPSKQATVKEPKVEPEAKVAA	300
QY	301	PEAKAEKPAVRPEPVPAAANTAASETAAESAPOEAAASALDPTDSTGNVSEPEVQVSAE	360
DB	301	PEAKAEKPAVRPEPVPAAANTAASETAAESAPOEAAASALDPTDSTGNVSEPEVQVSAE	360
QY	361	BETESGLPGSGYTL LLAGCGAALIALLLIRLAQSKRAARTTEESYPEPEPDLDDAADGGI	420
DB	361	BETESGLPGSGYTL LLAGCGAALIALLLIRLAQSKRAARTTEESYPEPEPDLDDAADGGI	420
QY	421	EITFAEVTPTAPEPAPKNDVNDT LALDGESEELSAKQTFDVTDTPSNRIDLDLDFDSL	480
DB	421	EITFAEVTPTAPEPAPKNDVNDT LALDGESEELSAKQTFDVTDTPSNRIDLDLDFDSL	480
QY	481	AAONGTILSALTCDEETQKRADNNAIESTSVTEPETFNPNVPEIVIDTPEPESVAQ	540
DB	481	AAONGTILSALTCDEETQKRADNNAIESTSVTEPETFNPNVPEIVIDTPEPESVAQ	540
QY	541	TAENKPEITVDITDFSDMLPSNNHIGTEETASAKPSPGLAGFLKASSPTEILEKTVAEVQ	600
DB	541	TAENKPEITVDITDFSDMLPSNNHIGTEETASAKPSPGLAGFLKASSPTEILEKTVAEVQ	600
QY	601	TPEELHDFLKVETDVAETAPEPTDFPNAADDLSALQAPASVPEENITETVAETPDF	660
DB	601	TPEELHDFLKVETDVAETAPEPTDFPNAADDLSALQAPASVPEENITETVAETPDF	660
QY	661	NATADDLSALQSEVPVAVENNAEIVADDLSALQAPAEVPEENVETVAETSDPHTA	720
DB	661	NATADDLSALQSEVPVAVENNAEIVADDLSALQAPAEVPEENVETVAETSDPHTA	720
QY	721	ADDLSSALQPAEVPVAVENVTKTVAEIPDFNATADDLSALQSEVPVAVENNAEITLET	780
DB	721	ADDLSSALQPAEVPVAVENVTKTVAEIPDFNATADDLSALQSEVPVAVENNAEITLET	780
QY	781	PDNSTSEADALPDLFKDGEETVDWISYILSEENI PNNADTSPFSESVGSDPASEAKYDLA	840
DB	781	PDNSTSEADALPDLFKDGEETVDWISYILSEENI PNNADTSPFSESVGSDPASEAKYDLA	840
QY	841	EMVLEIGDRDAAAEFTVOKLEEAEGDVLKPAQALAOELGI	880
DB	841	EMVLEIGDRDAAAEFTVOKLEEAEGDVLKPAQALAOELGI	880

RESULT 2
ABU37449
ID ABU37449 standard; protein; 801 AA.
XX
AC ABU37449;

23-OCT-2003 (revised)
19-JUN-2003 (first entry)
Protein encoded by Prokaryotic essential gene #22976.
Antisense; prokaryotic essential gene; cell proliferation; drug design.
Neisseria gonorrhoeae.
W0200277183-A2.
03-OCT-2002.
21-MAR-2002; 2002WO-US0009107.
21-MAR-2001; 2001US-00815242.
06-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-0362699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
N-PSDB; ACA41319.
New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 65373; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits cell
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
the target prokaryotic essential genes. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
standardise OS field)
Sequence 801 AA;
Query Match 78.0%; Score 3434; DB 6; Length 801;
Best Local Similarity 77.9%; Pred No. 1.1e-196;

RESULT 3
ABU37647
ID ABU
XX
AC ABU
XX

Db 1 NTAASETAESAQAASALDPTDGTGNAVSEPVQVSAEETESGLFDGLFGGSYTL 60
 QY 375 LLAGGGAAALALLLLHLAGSKARRTEESVPEEPDLDDAADDGIBITTAETVETATPE 434
 Db 61 LLAGGGAAALALLLLHLAGSKARRTEESVPEEPDLDDAADDGIBITTAETVETATPE 120
 QY 435 PAPKNDVNDTALDGESEELSAKOTFDVETDTPSNRIDLDLDFSLAAQNGILSGALTQD 494
 Db 121 PAPKNDVNDTALDGESEELSAKOTFDVETDTPSNRIDLDLDFSLAAQNGILSGALTQD 180
 QY 495 EETOKRADADWNALESSTSVPEPEFNPYNPVEIVIDTPPEPSVAQTAENKPKETVDTDFS 554
 Db 181 EETOKRADADWNALESSTSVPEPEFNPYNPVEIVIDTPPEPSVAQTAENKPKETVDTDFS 240
 QY 555 DNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYET 614
 Db 241 DNLPSNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYET 300
 QY 615 DAVAEAPETPDFNAADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPS 674
 Db 301 DAVAEAPETPDFNAADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPS 360
 QY 675 EVPAVEENAAEIVADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPAEVP 734
 Db 361 EVPAVEENAAEIVADDLSALLQPAEAPSVENITETVAEIPDFNATADDLSALLQPSVP 420
 QY 735 AVENVTVAEIPDFNATADDLSALLQPSVPAVEENAAETITETPDSTSEADALPDF 794
 Db 421 AVENNAETV-----ADDLSALLQPAEAPSVENITETVAEIPDFNATADDLSALLQPSVP 471
 QY 795 LKDGEEETVDWSIYLSSEENIPNADTSF 822
 Db 472 LKDGEEETVDWSIYLSSEENIPNADTSF 499

RESULT 4
 ABP80077
 ID ABP80077 standard; protein; 244 AA.
 AC ABP80077;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 6684.
 XX Antibacterial; infection; vaccine; gene therapy.
 XX Neisseria gonorrhoeae.
 XX WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB002069.
 XX 12-FEB-2001; 2001GB-00003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Massignani V, Monaci E;
 XX WPI: 2003-058415/05.
 XX N-PSDB; AB241047.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection.
 XX Disclosure; Page 674; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 XX Also disclosed are the nucleic acid molecules encoding the proteins and
 XX antibodies that specifically bind to the proteins. The composition

comprising the protein, nucleic acid or antibody is useful for the
 manufacture of a medicament for treating or preventing N. gonorrhoeae
 infection, this may be in the form of a vaccine or gene therapy.
 Sequences given in records ABP76736-ABP81046 represent nucleic acid
 molecules of the invention
 Sequence 244 AA;
 Query Match 25.5%; Score 1122.5; DB 6; Length 244;
 Best Local Similarity 64.6%; Pred. No. 2.9e-59;
 Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;
 QY 514 VYBPETNPYNPVEIVIDTPPEPSVAQTAENKPKETVDTDFSNDLPSNNHIGTEETASAKP 573
 Db 1 VYBPETNPYNPVEIVIDTPPEPSVAQTAENKPKETVDTDFSNDLPSNNHIGTEETASAKP 60
 QY 574 ASPGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYETDAVAETAPETPDFNAADD 633
 Db 61 AAPGLAGFLKASSPETTILEKTVAEVQTPPELHDFLKVYETGAETAPETPDFNAADD 120
 QY 634 LSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPSVPAVEENAAEIVADDLSA 693
 Db 121 LSALLQPAEA----- 130
 QY 694 LLQPAEAPSVENITETVAETSDPHTAADDLSALLQPAEVPVAVENVTKTVAEIPDFNAT 753
 Db 131 ----- 130
 QY 754 ADDLSALLQPSVPAVEENAAETITETPDSTSEADALPDFLKDGEETVDWSIYLSSEEN 813
 Db 131 -----PAVEENAAETITETPDSTSEADALPDFLKDGEETVDWSIYLSSEEN 177
 QY 814 IPNNADTSPPSESVGSDAPSEAKYDIAEMYLEIGORDAAETVQKLLSEAGDVLKRAQA 873
 Db 178 IPNNADTSPPSESVGSDAPSEAKYDIAEMYLEIGORDAAETVQKLLSEAGDVLKRAQA 237
 QY 874 LAQELGI 880
 Db 238 LAQELGI 244
 RESULT 5
 ABP78566
 ID ABP78566 standard; protein; 244 AA.
 AC ABP78566;
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae amino acid sequence SEQ ID 3662.
 XX Antibacterial; infection; vaccine; gene therapy.
 XX Neisseria gonorrhoeae.
 XX WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB002069.
 XX 12-FEB-2001; 2001GB-00003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Massignani V, Monaci E;
 XX WPI: 2003-058415/05.
 XX N-PSDB; AB239536.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection.

PS Disclosure; Page 453; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX Sequence 244 AA;

XX Query Match 25.5%; Score 1122.5; DB 6; Length 244;

XX Best Local Similarity 64.6%; Pred. No. 2.9e-59;

XX Matches 237; Conservative 2; Mismatches 5; Indels 123; Gaps 1;

QY 514 VYEPETNPYNPVEIVIDTPEPSVAQTAEKNKPTVDTDFSDNLPNNHIGTEETASAKP 573

DB 1 VYEPETNPYNPVEIVIDTPEPSVAQTAEKNKPTVDTDFYNNLFSNNHIGTEETASAKP 60

QY 574 ASPGSLAGFLKASSPETILEKTVAEVQTPBELHDFLKVYETDAVETAPETPDFAAAD 633

DB 61 AAPGSLAGFLKASSPETILEKTVAEVQTPBELHDFLKVYETGAVETAPETPDFAAAD 120

QY 634 LSAQLQPAEAPSEVENTETVETPDNPATADDLSALLQPSVEPVAEENAAEIVADDLSA 693

DB 121 LSAQLQPAEA----- 130

QY 694 LLQPAEAPAEVENTETVETVETVETVETVETVETVETVETVETVETVETVETVETVET 753

DB 131 ----- 130

QY 754 ADDLSALLQPSVEPVAEENAAEITLTPDSNTSEADALPDFKDGRETVDMWISYI SEEN 813

DB 131 -----PAVEENAAEITLTPDSNTSEADALPDFKDGRETVDMWISYI SEEN 177

QY 814 IPNADTSPFSEVSGDAPSEAKYDLAEMYLEIGDRDAAAEITVQKLEAEGDVLKRAQA 873

DB 178 IPNADTSPFSEVSGDAPSEAKYDLAEMYLEIGDRDAAAEITVQKLEAEGDVLKRAQA 237

QY 874 LAQELGI 880

DB 238 LAQELGI 244

RESULT 6

ABP78509

ID ABP78509 standard; protein; 275 AA.

AC ABP78509;

XX 07-MAR-2003 (first entry)

DE *N. gonorrhoeae* amino acid sequence SEQ ID 3548.

XX Antibacterial; infection; vaccine; gene therapy.

XX *Neisseria gonorrhoeae*.

OS WO200279243-A2.

PN 10-OCT-2002.

PD 12-FEB-2002; 2002WO-IB002069.

PF 12-FEB-2001; 2001GB-00003424.

PR (CHIR-) CHIRON SPA.

PA Fontana MR, Pizza M, Massignani V, Monaci E;

PI WPI; 2003-058415/05.

DR

DR N-PSDB; ABZ39479.

XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing *N. gonorrhoeae* infection.

XX Disclosure; Page 447; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX Sequence 275 AA;

XX Query Match 23.3%; Score 1025; DB 6; Length 275;

XX Best Local Similarity 83.4%; Pred. No. 2.2e-53;

XX Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;

QY 107 LGDKAVIAVSSEQAVRDPVLVFRIGAGAOVREYTAIILDPVGYSPKTKSALSDGKTHRKTA 166

DB 1 LDGKAVIAVSSEQAVRDPVLVFRIGAGAOVREYTAIILDPVGYSPKTKSALSDGKTHRKTA 60

QY 167 PTAESQENQNAKALRTDKKDSANAAVKPAINGKTHTVRKGETVKQIAAAIRPKHLTLEQ 226

DB 61 PKAESQENQNAKALRTDKKDSANAAVKPAINGKTHTVRKGETVKQIAAAIRPKHLTLEQ 120

QY 227 VADALLKANPNVSAHGRLRAGSVLHPIPNLR-----IKAEQPKQTAETASMP 278

DB 121 VADVLLKANPNVSAHGRLRAGSVLHPIPNLR-----IKAEQPKQTAETASMP 180

QY 279 SEPSKQAT-----VEKPVKEPAKVAAPAEKPAKPAVRP----- 313

DB 181 SEPSKQATVEKPIKPVKEPAKVAAPAEKPAKPAVRP----- 313

QY 314 -PVPAANTAASETAASAP 331

DB 241 GPVPAANTAASETAASAP 259

RESULT 7

ABP79684

ID ABP79684 standard; protein; 275 AA.

XX ABP79684;

XX 07-MAR-2003 (first entry)

DE *N. gonorrhoeae* amino acid sequence SEQ ID 5898.

XX Antibacterial; infection; vaccine; gene therapy.

XX *Neisseria gonorrhoeae*.

OS WO200279243-A2.

PN 10-OCT-2002.

PD 12-FEB-2002; 2002WO-IB002069.

PF 12-FEB-2001; 2001GB-00003424.

PR (CHIR-) CHIRON SPA.

PA Fontana MR, Pizza M, Massignani V, Monaci E;

PI WPI; 2003-058415/05.

DR N-PSDB; ABZ40654.

XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT

PT medicament for treating or preventing N. gonorrhoeae infection.

XX Disclosure; Page 619; 815pp; English.
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention
 XX Sequence 275 AA;
 SQ

Query Match 23.3%; Score 1025; DB 6; Length 275;
 Best Local Similarity 83.4%; Pred. No. 2.2e-53;
 Matches 216; Conservative 3; Mismatches 6; Indels 34; Gaps 3;
 QY 107 LGDKAVTAVSFEQVRDPVLFRIGAGAOVREYTAILDPVGYSPKTKSALSDGKTHRKA 166
 DB 1 LDGKAVTAVSQAQVRDPVLFRIGAGAOVREYTAILDPVGYSPKTKSALSDGKTHRKA 60
 QY 167 PTASQENQNAKLRKTDKDSANAAPKPYNGKTHVVRKGETVKQIAAARPKHLTLEQ 226
 DB 61 PKAESQENQNAKLRKTDKDSANSAPKPYNGKTHVVRKGETVKQIAAARPKHLTLEQ 120
 QY 227 VADALLKANPNVSAHGRURAGSVLHLPNLR-----IKAEQPKPOTAKPKAETASMP 278
 DB 121 VADVLKANPNVSAHGRURAGSVLHLPNLR-----IKAEQPKPOTAKPKAETASMP 180
 QY 279 SEPSQKAT-----VKPKVEKPEAKVAAPKAEKPAVRPE----- 313
 DB 181 SEPSQKATVEKPKVEKPEKPEAKVAAPKAEKPAVRPEKPAVSETPASATERQP 240
 QY 314 -PVPAANTAASETAESAP 331
 DB 241 GPVPAANTAASETAESAP 259

RESULT 8
 ABP78538
 ID ABP78538 standard; protein; 253 AA.
 XX AC ABP78538;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 3606.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB002069.
 XX PR 12-FEB-2001; 2001GB-00003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX DR N-PSDB; ABZ39508.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection.
 XX PS Disclosure; Page 450; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention
 XX Sequence 253 AA;
 SQ

Query Match 18.7%; Score 821.5; DB 6; Length 253;
 Best Local Similarity 66.8%; Pred. No. 2.9e-41;
 Matches 181; Conservative 11; Mismatches 40; Indels 39; Gaps 5;
 QY 255 LNRKAEQPKPOTAKPKAETASMPSEPSKQATVEKPKVEKPEAKVAAPKAEKPAVRPEP 314
 DB 1 LKNLKQKLP-----KQKRNRPD-PNPNFQCPKLRRLRRQVNPDL 42
 QY 315 VPAANTAASETAESAPQEAASAIIDTPTDETGNVS-----BPVEQVS 358
 DB 43 YPLQILPHRKPLPNPPPPQEAASAIIDTPTDETGNVS-----BPVEQVS 102
 QY 359 ABEETES-----GLFGGSYTLILLAGGGAALIA-LLLLLRLAQSKRARTTEESVPPEEPDL 413
 DB 103 ABEETESGLFDGLFGGSYTLILLAGGGAALIA-LLLLLRLAQSKRARTTEESVPPEEPDL 162
 QY 414 DAADDGIEITFAEVETPATPEAPKQVNDTLALDGESEBELSAKQTFDVTDTSPNRID 473
 DB 163 DAADDGKITFAEVETPATPEAPKQVNDTLALDGESEBELSAKQTFDVTDTSPNRID 222
 QY 474 LDFDLSAAQNGILSGALTQDEETOKRADAD 504
 DB 223 LDFDLSAAQNGILSGALTQDEETOKRADAD 253

RESULT 9
 ABP79950
 ID ABP79950 standard; protein; 132 AA.
 XX AC ABP79950;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 6430.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB002069.
 XX PR 12-FEB-2001; 2001GB-00003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX DR N-PSDB; ABZ40920.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection.
 XX PS Disclosure; Page 656; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and

antisen nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the gene; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at [ftp.wipo.int/pub/published/pct](http://wipo.int/pub/published/pct) sequences

582 LKLMVEYAEVMDREGFARQENELREIGGAQ-PQVEQLKSRV-----PAMVAVAAVA 631
 535 PESVAQTAENKPTVDTDFSDNLPSSNNHIGTEBTASAKP-ASPSGLAGFLKASSPETILE 593
 632 GLAGAKLAQDELDSFSLD-DLSLDDSGH-----AAKPDAAAGQDLDADFSLDDLLGGD 683
 594 KTVAEVQTPPE-ELHDFLKVVETDAVATAPETPDFAAADDLSALLQPAEAPSVEENITE 652
 684 DVQADLKSDGALDLDLSDSLDLAATPADKP-----VDDLDFGLDFAG----- 728
 653 TVAETPDFAADLSDALLQSEVPAVEENAEIVADDLSALLQ---PAPAPAVEENITE 709
 729 -LAETPS-QPKHDLGDFSLDLPADP---EDKLSD---DFFLLSNDVEVPAAPADNEFTLD 781
 710 TVAETSPHTAAD--DLSALLQPAEVAEENVTVAETPDFAADLSDALLQ---PAPAPAVEENITE 767
 782 TEAAEELPALSFPDFSLDLPADP---EDKLSD---DFFLLSNDVEVPAAPADNEFTLD 833
 768 AVENAAEITLETDSNT---SEADALPDFLKDGEEETVDSIYLSSEENIPNNADTSPPS 824
 834 -----NLDPEKSTPSPSADAAVASALDGD-----ADDDDFD 866
 825 EVSGSDAPSEAKYDLAEMLYRIGDRDAAAEVQKLLAEAGDVLKRAQA 873
 867 LSGADERAAT--KLDLARAYIDMGDSGARDILDEVIL--AEGNDSQQAFA 911

RESULT 11
 ID ABJ26247
 XX AC ABJ26247
 XX DT 17-APR-2003 (first entry)
 XX DE Aspergillus fumigatus essential gene protein #905.
 XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX KW cancer; contamination; biofilm; antibody; immune response.
 XX OS Aspergillus fumigatus.
 XX PN WO200286090-A2.
 XX PD 31-OCT-2002.
 XX PF 23-APR-2002; 2002WO-US013142.
 XX PR 23-APR-2001; 2001US-0285697P.
 XX PR 27-APR-2001; 2001US-0287066P.
 XX PR 05-JUN-2001; 2001US-0295890P.
 XX PR 09-JUL-2001; 2001US-0303899P.
 XX PR 31-AUG-2001; 2001US-0316362P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX DR New purified or isolated nucleic acids of essential genes of Aspergillus
 XX PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 XX PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX PS Disclosure; Page; 175pp; English.
 XX The invention relates to novel purified or isolated nucleic acids of
 XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 XX the invention are used to treat or prevent infections by a pathogenic
 XX organism such as A. fumigatus, to treat a non-infectious disease in a
 XX subject (e.g. cancer), to prevent or contain contamination of an object
 XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
 XX biofilm comprising A. fumigatus. The polynucleotides are useful for

expressing recombinant protein for characterisation, screening or
 therapeutic use, as markers for host tissues in which the pathogenic
 organisms invade or reside, for comparing with the DNA sequence of A.
 fumigatus to identify duplicated genes or paralogues having the same or
 similar biochemical activity and/or function, for comparing with DNA
 sequences of other related or distant pathogenic organisms to identify
 potential orthologous essential or virulence genes, for selecting and
 making oligomers for attachment to a nucleic acid array for examination
 of expression patterns, for raising anti-protein antibodies, as an
 antigen to raise anti-DNA antibodies or to elicit another immune
 response, and for identifying polynucleotides encoding the other protein
 with which binding occurs or to identify inhibitors of the binding
 interaction. The polypeptides may be used to raise antibodies or to
 elicit immune response, as a reagent in assays designed to quantitatively
 determine levels of the protein in biological fluids, as a marker for
 host tissues in which pathogenic organism invade or reside, and to
 isolate correlative receptors or ligands in the case of virulence
 factors. This sequence represents a protein of one of the essential genes
 of Aspergillus fumigatus of the invention

Sequence 1095 AA;
 Query Match 8.1%; Score 358; DB 6; Length 1095;
 Best Local Similarity 22.1%; Pred. No. 9.5e-13;
 Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;

QY 38 NRQIKLIAASVAAAFQAHAGLGLINIQSLNLPFGSGITVTGEEAKALGGSVTVSE 97
 DB 91 NNILTVGSSSIAAAPGAENH-----VEKETETPLENGASEKVEETAKEPGVQSTTTTE 144
 QY 98 KGLTAKVHKLGDKAVIA-----VSSFOAVRD--PVLVFRIGAGAOVRE--YTALIDPVGYS 149
 DB 145 AVKDEVPEKTDSDSVVVEKDLHVESEPAEQVVKPLVSEPKSPDIHEDAEQTPASNAAE 204
 QY 150 PKTKSALS-----DGKTHRTKPTAPAESQENQNAKALRKTDKOSANAARKPAYNGK 200
 DB 205 PKTNGTHSSEIQSGQPDPAVTEKVEPVEEKAADLTITTKVANKPPEIPSTSLAEKPE 264
 QY 201 THTVRKGE-----TVKQIAAIRPKHL---TLEQVADALLKANPNVSAHGLRL 245
 DB 265 TKEVEKAEDQKLQDLEIVPAIVEEKEAASEVPESTLEKSGEAAAPQ--TVTEVPLV 323
 QY 246 AGSVLHINLNRIKAEQ--PKPQTA--KPKAETAS---MPSEP-----SKQATV----- 287
 DB 324 DSKPVLSAESTKAEKAPTMTTDEAPLAEKSEKTVPAEIAEKVQATIGDEPLSRN 383
 QY 288 ---EKPVKPK-EAKVAAPAEKAEKPAVRPEPVPAAATAAETAASASQAEAAASAI-DTP 342
 DB 384 NVVEEQPKQSVKAEKAADEAVLEEP-----IKESAPBEIPEDSRNAVVDAP 430
 QY 343 TDETG-----NAVSEPVQVSAEETESGLFGGSYTLILLAGGGAALLALLLLRL 392
 DB 431 VTESSTTKVTVTAPDAEKEPAKESAKEPIQEEIPEASEKEVAEE-----TPV 479
 QY 393 AQSKRAARTEESV--PEPEPDLDDAADDGI-----EITPA-EVETPATPEPAP 437
 DB 480 TESSTTEKVEVTDALAEKEPALEESTKEPIHBEQTIETVTAETVAAKEPEAQATEPVA 539
 QY 438 KNDVNDTLALDGESBE---ELSAKQTFDVT-----DTPSNRIDLDLDSLAQAONGI 486
 DB 540 NESINK--EVEATTEPTAGQESAKEPISNETAIRVTOADESTKELSTGF---AAADIT 594
 QY 487 LSGALTQDDET-----QGRADADWNAIESTD-----512
 DB 595 GKGAETHTEVSTAAGPAKEFVNDASAVKSAPTETVQETKDDSAQVVGKAEAPQIAEE 654
 QY 513 -SVYEPETNPVNPVEIVIDTPEP-----ESVAQTAEKPKETVD 550
 DB 655 AAVEETSAANPEAPTEESTTAEPAKEPGESEETPCQTLPAEESKEATEAAEAAKEBSTIE 714
 QY 551 TDFSDNL-----PSNNHIGTEBTASAKPASPGLAGFLKASSPETILEKTVAEVQTPPE 604
 DB 715 T-VSEPLEADVKPEAQDDPATETLTEDKPVQBSVAE--AAPKPESEVPVSEVSGTAKTS 771

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:15:51 ; Search time 20 Seconds
(without alignments)
2271.543 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRCPMMTKFTDCT.....EBAEGDVLKRAQALAEIGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
1: /cgn2.6/prodata/2/iaa/5A COMB.pdp.*
2: /cgn2.6/prodata/2/iaa/5B COMB.pdp.*
3: /cgn2.6/prodata/2/iaa/6A COMB.pdp.*
4: /cgn2.6/prodata/2/iaa/6B COMB.pdp.*
5: /cgn2.6/prodata/2/iaa/PCUS COMB.pdp.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	398.5	9.0	975	US-09-252-991A-28749	Sequence 28749, A
2	315.5	7.2	1786	US-08-973-462-8	Sequence 8, Appli
3	289	6.6	8991	US-08-714-741-32	Sequence 32, Appl
4	276.5	6.3	2468	US-09-976-594-726	Sequence 726, App
5	269	6.1	540	US-08-973-462-22	Sequence 22, Appl
6	267	6.1	688	US-09-141-047-8	Sequence 8, Appli
7	250.5	5.7	630	US-08-973-462-9	Sequence 9, Appli
8	236	5.4	703	US-09-252-991A-17865	Sequence 17865, A
9	236	5.4	1596	US-08-978-277A-4	Sequence 4, Appli
10	235	5.3	1346	US-08-635-121-2	Sequence 2, Appli
11	235	5.3	1346	US-08-978-277A-2	Sequence 2, Appli
12	235	5.3	2137	US-09-134-001C-4463	Sequence 4463, Ap
13	233.5	5.3	2482	US-09-252-991A-16967	Sequence 16967, A
14	232.5	5.3	1848	US-08-296-791-6	Sequence 6, Appli
15	232.5	5.3	1848	US-09-839-996-6	Sequence 6, Appli
16	232.5	5.3	1848	US-10-080-505-6	Sequence 6, Appli
17	232.5	5.3	1848	PCT-US95-10661A-6	Sequence 6, Appli
18	232.5	5.3	2315	US-09-543-681A-5434	Sequence 5434, Ap
19	230.5	5.2	1018	US-08-072-610-2	Sequence 2, Appli
20	230.5	5.2	1018	US-08-719-822B-2	Sequence 2, Appli
21	230.5	5.2	1018	US-09-092-458-2	Sequence 2, Appli
22	217	4.9	1231	US-08-904-263A-4	Sequence 4, Appli
23	217	4.9	1231	US-09-434-123A-4	Sequence 4, Appli
24	216.5	4.9	1075	US-09-252-991A-18387	Sequence 18387, A
25	214	4.9	928	US-09-134-000C-6590	Sequence 6590, Ap
26	213.5	4.8	1601	US-09-345-473E-40	Sequence 40, Appl
27	213	4.8	1702	US-08-296-791-5	Sequence 5, Appli

28	213	4.8	1702	4	US-09-839-996-5	Sequence 5, Appli
29	213	4.8	1702	4	US-10-080-505-5	Sequence 5, Appli
30	213	4.8	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
31	212.5	4.8	1404	4	US-10-164-595-78	Sequence 78, Appl
32	212.5	4.8	2756	1	US-08-375-709-11	Sequence 11, Appl
33	212.5	4.8	2756	1	US-08-752-929-11	Sequence 11, Appl
34	212.5	4.8	2756	3	US-09-090-793-7	Sequence 7, Appli
35	212.5	4.8	2756	4	US-09-231-899-7	Sequence 7, Appli
36	211.5	4.8	3788	4	US-09-336-447A-76	Sequence 76, Appl
37	210.5	4.8	1140	4	US-07-757-022B-104	Sequence 104, App
38	210.5	4.8	1363	4	US-07-757-022B-52	Sequence 52, Appl
39	210.5	4.8	1404	4	US-07-757-022B-2	Sequence 2, Appli
40	210.5	4.8	1404	4	US-07-757-022B-62	Sequence 62, Appl
41	210	4.8	941	4	US-07-757-022B-14	Sequence 14, Appl
42	210	4.8	1022	4	US-07-757-022B-84	Sequence 84, Appl
43	210	4.8	1038	4	US-07-757-022B-74	Sequence 74, Appl
44	210	4.8	1049	4	US-07-757-022B-58	Sequence 58, Appl
45	210	4.8	1270	4	US-07-757-022B-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-28749
; Sequence 28749, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28749
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28749

Query Match 9.0%; Score 398.5; DB 4; Length 975;

Best Local Similarity 23.1%; Pred. No. 1.8e-20;

Matches 240; Conservative 136; Mismatches 378; Indels 285; Gaps 44;

QY	14	TKFTDC---TRNRIOPTTHRGVILKNNRQIKLIASVAVASFOAHA-GLGLNIOQNL	69
Db	35	TASTDALPLVKEKILKP--RDYTWRLTLVRAIAAASVLTSGMAHGLGLGELITLKSAL	92
QY	70	DBPFGSITVTGEEAKALGGGSVTVS-----BKGLTAKVHKL-----GDKA	111
Db	93	NQPLDAEIELL--EVRD-LGSEVIPSLASPEFSKAGVDRLYLTLDKFTPVKENGKS	149
QY	112	VIATVSEQAVRDPVLVFRIGA---GAQVREYTAILDPVGYSPKTKSALSDGKTHKRTAP	167
Db	150	VIRVTSKPEQVQSPYFLNVLQVLPNGRLIREYTLVLDPLPSYQAAASAPAPV---SAP	206
QY	168	TAESQENQAKALRTDKDSDANAIVK---PAYNGKTHVTKGETVKQIAAIRPKHLTL	224
Db	207	RATG-----APRAPQAPAPVPTAPGSDTYRTV-SNDTLWEIAQRNTRDRSV	254
QY	225	EQVADALLKANPNVSAHG---RLRAGSVLHINLRKAEQPKPOTAKPKAETAS-----	276
Db	255	PQAMLAFOELNFGAFVDGNINELKSGQVLRITPEQOMLERSREALSQVQAQNSWRGR	314
QY	277	-----MPSEPSK-QAT-----	286
Db	315	NPAAGSAGARQIDA*QRNAAGSAPSKVDATDNLRLVSGEGKASKGADKGGKSKALADT	374

QY 287 -----VEKPVKPEAKVA----- 299
Db 375 LAVTKESLDSRRNEELQSRMQLQSLQDLKQLKLIQKDAQLAKQQLGAEGQAQP 434
QY 300 -----APAKAEKPAVRPEVPVPAANTAASETAAESAPOEAAAASAIPTDE--TGNA 349
Db 435 NAALFDASQPNAAQAQAFQGTAAAAAPTTPAGEAFAAPAQAQPPVAPPAEKKPPAPA 494
QY 350 VSEPEVQVSAEETESGLFGGSYTLILAG-----GGAALIALLLRLAQSRRARTE 402
Db 495 VPAPVQAQAPAPSEI-----DELLANPLWLAVIGGSALLLVLLMILSRNAQEK 549
QY 403 ESEV-----EEFPDLDDAADDGI-EITFAEVETPATPEPAPKNDVNDTLALDGESEEE 454
Db 550 EEVQAFADAGEEQEDALDLKGDFDDLTLDEPE-PQVAAAAPQVEKTTAQTSDALGEAD 608
QY 455 L-----SAKOTFVETTPSNRIDLDFDSLAA-AQNGILSGALTQDEETOKRADAD 504
Db 609 IYIAYGRFNOAAELLQNAIYDEP-QRTDLRLKMEVVAEMGDREGFAQENELREIGGAQ 667
QY 505 WNAIESTDSVVEPETENPNVPEIVIDTPPEPVSAQTAENKPEVDTDFSNLPSNNHIG 564
Db 668 -PQVEQLKSY-----PAMVAVAAVAGLAGKLAQDELSFSLD-DLSLDSGH-- 714
QY 565 TEETASAKP-ASPSGLAGFLKASSPETTILEKTVAEVQTPPE-ELHDFLKVYETDAVAETAP 622
Db 715 -----AAKPDAAQQLDDAFDLSLDLDGDDVQADLKSDSGALDDBLTLDSLDLAASTAA 769
QY 623 ETPFNAAAADLSALLQPAEAPSVENITETVAETPDNATADLSALLQSEVPAPVEN 682
Db 770 DKP-----VDLDFGLDFAE-----LAETPS-QPKHDDLGLDFSLDLADP--BDK 810
QY 683 AAEIVADLSALLQ-----PAEAPAVENVTETVAETSDPHTAAD--DLSALLQPAEVPAVE 737
Db 811 LSD-----DFLLSLNDEVPAAPADNEFTLDEAAEPAISLPDDFDLSLADEPTEPAPE 867
QY 738 ENVTKTVAEIPDFNATADLSALLQSEVPAPVEENAAITLETPTSNT---SEADALPDF 794
Db 868 KGEDSFAQLDEVSQAQLDELAS-----NLDEPKSATPSFAEDAAVAS 910
QY 795 LKDGEEVVDWSIYLSEENIPNNAATSPSBSVGDAPSEAKYDLAEMYLEIGDRDAAR 854
Db 911 ALDGD-----ADDDFDLFGADEAAT--KLIDARAYIDMGDSEGARD 950
QY 855 TVQKLEEAEGDVLKRAQA 873
Db 951 ILDEVL--AEGNDSQAQA 967

RESULT 2

US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0650-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 7.2%; Score 315.5; DB 3; Length 1786;
Best Local Similarity 20.2%; Pred. No. 5e-14;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIQNLDEPFGSGITVTGEEAKALLGGSVTVSEKGLTAKVHKLGDKAV---IAVSSEQA 120
Db 242 NVEENBENDDGSVASSVES-----IASSVDESIDSIEENVAPTVEEI 286
QY 121 VRDVLVFRICAGAVREYTAILDPGVSPKTSALSADGKTHRTAPTAAES--QENQNAK 178
Db 287 VAPSV-----ESVAPSVESVEENVEES 310
QY 179 ALRXTDKKDSANAAVKPAYNGKTHVRKGE-----TVKQIAAAIRPKHLTLEQ--VADALL 232
Db 311 VAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRRLRAGSVLHLPNLRIRKAEQPKQTAKPKAETASMPSEPSKQATVVKPV 291
Db 365 ESVAPSVESVE-----ENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAAPR-----AKAEKPAVRPEVPVPAANTAASETAAESAPOEAAAASAIPTDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSVESVEENVEESVAENVEESVAENV 475
QY 347 GNAYSEPVQVSAEETESGLFGGSYTLILAGGGAALIALLLRLAQSRRARTEESPV 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDADAADDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKOTFVET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSVESVEE-----NVVEE 557
QY 466 DTPSNRIDLDPDSLAAAQNGILSGALTQDEETOKRADAWNIAIESTDSVYEPFENPNYP 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPPEPVSAQTAENK- PETVDTDFSDNLPSSNNHIGTEETASAKPSPGLAGFL 583
Db 602 TVEEIVAPSVSVESVAPSVESVEENVEESVAENVEESVAENVEESV----- 647
QY 584 KASSPETTILEKTVAEVQTPPEELHDFLKVYETDAVAETAPETPDFNAAAADLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVSVESVAPSVESVEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVAETPDFNATADLSALLQSEVPAPVEENAAITVADLSALLQPAEAPAV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVSVESVAPSVESV 759
QY 704 EENVETVAETSDPHTAADLSALLQPAEVPAPVEENVTKTVAEIPDFNATADLSALLQ 763
Db 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSVEE 805
QY 764 SEVPAPVEENAAITLETPTSNTSEADALPDFLKCGEEETVDWSIYLSEENIPNNAATSP 823
Db 806 SVAPSVESVA-----ENVATNLSD--NLLSNLGGTETETIKDSILNEIEEVKENVVTIT- 859
QY 824 SESVGSAPSEAKYDLAEMYLEIGDRDAAAATVQKLEEAEGDVLKRA 871
Db 860 LENV-EETTAESVTTFSNILEEIQENTITNDTIEBKLEELHENVLSAA 906

RESULT 3

US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca
 APPLICANT: Brooks-Walter, Alexis
 TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
 TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
 TITLE OF INVENTION: PORTIONS AND PRODUCTS
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtis, Morris & Safford, P.C.
 STREET: 530 Fifth Avenue
 CITY: New York
 STATE: New York
 COUNTRY: U.S.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/714,741
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Frommer Esq., William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454312-2460
 TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8991 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: amino acid
 US-08-714-741-32

Query Match 6.6%; Score 289; DB 4; Length 8991;
 Best Local Similarity 21.6%; Pred. No. 4.8e-11;
 Matches 198; Conservative 92; Mismatches 313; Indels 312; Gaps 34;
 QY 146 VGYSPKTSALSGKTHRTAFAESENONAKALRT--DKKDSANAAVKPAYNGKTH 203
 DB 7741 VEYFKT-----DAEQTEQYLAARLADKKALEKTEADLKKAENEKPAETPAPA 7795
 QY 204 VRKGETVKOLAAAIRK-----HLTLEQVADALLKANPNVSAHGRLAGSVLHLPN 256
 DB 7796 PKPEQAEQPKAPAPQAPAPKPEKTDQQA-----EDYARRSEEEYNRLPQQ 7846
 QY 257 RIAEQPKP-----QAKPKAETASMPSEPSKQATVEKPKPEAKVAAP 301
 DB 7847 PKPAEKAPAPKPEQVPVPAEXPENPAPAPKPAKAPQPLKPEEPAEQPKP-EKPEEPAGQP 7905
 QY 302 E-----AKAEKPAVRP-----EPVPAANT-----320
 DB 7906 EPEKDDQAGEDIARRSGEYNRFPQQPPKAEKAPAPKPEQVPVPAKTLKKAKLAG 7965
 QY 321 AASETAESA-----PQEAASADTPTDETGNAY 350
 DB 7966 AKSKAATKAAELEPELEKAELENLLSTLDPGKTQDELDEKAEAELENKVEALPNQV 8025
 QY 351 SEPVEQVSAEETESLFGSGVTLTLAGGGAALIALLLLRLLAQSKRARRTEESVPEEP 410
 DB 8026 SELEELSLELNLKDAETNNVEDYIKG-----LEAIAIKQAELEKTPKELDA 8075
 QY 411 DDDAAADDGIEITFAEVETPA-----TPEAPKNDVN-DTALDQ-----ESEER- 454
 DB 8076 ALNELGPDGDEBETPPPEAPAEQPKPEAEETPAPAPKPEKSADQAEEDYARRSEERY 8135
 QY 455 --LSAKQTFDVTDTPS-----NRIDLDPSL-----479
 DB 8136 NRLTQQQPKAEKAPAPAPKPEQAPAPKPSRGLATKKKLNLAERIELLLKGLLEPGL 8195

QY 480 --AAQNGILSGAL-----TQDETQKRADWNNAIESTSVSEVPEFENPNPVEIIVDT 532
 DB 8196 EKAGAGLGNLLSTLDPGKTQDELDEKAEAELEN-----KKVEAL-----8235
 QY 533 PEPESVAQTAENKPEVTDTDFSNLPSNNHIGTEETASAKPASPGLAGFLKASSPETIL 592
 DB 8236 --PNQVAELBEEL-----SKLEDNL-----KDAETNHVEDYIKGLEEAIA 8274
 QY 593 EKTVAEVQTPTEELHDFLKVYETDAVETAPETDFNAAADDLSALLQAPAPAPSVEENITE 652
 DB 8275 TKQAELEKTPKELDAALNELGPDGDEE---ETPAPEA-----PAPQPKPEKPAEE 8321
 QY 653 TVAETPDENATADDLSALLQSEVPVAVENAAEIVADDLSALLQ-----PAEAPAVE 704
 DB 8322 TPAPAPKPEKSADQ-----QAEDYARRSEEEYNRLTQQQPPKAEKAPAPAPK 8370
 QY 705 ENVTVAETSDPHTAADDLSALLQPAEVPVAVENVT-----KTVAEIPDFNA 752
 DB 8371 PE-----QAPAPKPKQKVNLENLLSTLDPGKTQDEL-DKGA 8407
 QY 753 TADDLSALLQ--PSEVPVAVENAAEITLETPTSNTSEADA--LPDFLKQGEETVDSIY 808
 DB 8408 AEALNKKVEALPNPVKLEEE-----LSPEDNLKDAETNHVEDYIKGLEEAIAIKQA 8462
 QY 809 LSEEN-----IPNNADTSFPSESVGSDAPSEAKYDLAEMYLEIGDRDAAAEVQ 857
 DB 8463 ELEETPOEVDALNDLVPDGGEEETPAPAPQDEPAPAPAPNAEQPAPAPKPEKSAD---8519
 QY 858 KLEEAEGDVLKRAQ 872
 DB 8520 ---QAEEDYARRSE 8531

RESULT 4

US-09-976-594-726
 ; Sequence 726, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 726
 ; LENGTH: 2468
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
 US-09-976-594-726

Query Match 6.3%; Score 276.5; DB 4; Length 2468;
 Best Local Similarity 20.1%; Pred. No. 5.6e-11;
 Matches 185; Conservative 125; Mismatches 333; Indels 277; Gaps 35;

QY 150 PKTKSALSDGKT--HRKTAPTAESENONAKALRTDKKDSANAAVKPAYNGKTHTVRK 207
 DB 656 PKPEVAKDEKTPPKKEEKPKKEEYKVEKKEIKKEEKKEPKKEVKEKTPPKPEVKEVKK 715
 QY 208 ETVKQIAAAAIR--PKHLTLEQVADALLKANPNVSAHGRLAGSVLHLPNLRKAKQPKQ 266
 DB 716 EEKKVKEEKEPKKEIKKLPDAAKSSSTP-----LSBAKPA 753
 QY 267 TAKPKAETASMPSEPSKQATVE-----KPKVEKPAKVAAPAEKAEKPAVPEPVPAN 319
 DB 754 ALKPKV-----PKKEESVKKDSVAAGKPKKEKIKKIKKEGKAAE-AVAAAVGTGAT 804

QY 320 TAASETAESA-----PQAAAAIAIDTPTDTCNAVSEPVQVSAEEETESGLFGGYSY 373
Db 805 TAAMAAAGTAAIGAPAKELAEKSLMSPEDLTQDFELKAEVDVTKDKP----- 856
QY 374 LLLAGGGAALIALLLALLLAQSKARTEESVPEEPDLD-----DAADDGIEIITFAEVET 429
Db 857 -----QLELIDEEKLLKETEPVEAVVIKEREVTKGPAESPDEGITTTEGEGEC 905
QY 430 PATP-----EPAPKNDVNTLAL-----DGE----- 450
Db 906 EQTPEELEPVKQGVDDIEKFEDEGAGFESESTGDYEEKAEETAEAEPEEDGSEHVCVS 965
QY 451 -----SEELSAKQTPDV-----ETDTSNRIDLDFOGLAAQNGILSCALTDQDEET 497
Db 966 ASKHSPTHEDESAKAEDAYIREKRESVASGDDRAEDMDZ--ALIEG-----BAQS :016
QY 498 QKRADADNNALESTDSVYEPETFPNPNPVEIVID-----TPESVVAQT 541
Db 1017 EERAEDEKADAREEYEPKMEAEYVMAVDKAAEAGAEQYGLTTPTKQLGAQS 1076
QY 542 AENKPEITDTPSDNLPNNH-----IGTEETASAKPASPSGLAGFLKA----- 585
Db 1077 PGRFPASSIHD--ETLPGGSESEATASDEENREDQPEFTATSGTQSTIIEISSEPTMD 1134
QY 586 --SSPETILEKTVAEVQTPPELHDFLKV--YETDAVAE---TAPETP----- 625
Db 1135 EMSTPRVMSDETNNETESPSQEFVNITKYESLSYQSEYKPADVTPLNGFSESGKTD 1194
QY 626 ----DFNAAA-----DLSAL-----LOPA 641
Db 1195 TDGKDYNASASTISPPSMEEBKFSRSLRDAYCSEVKASTTLDIKDSISAVSSKVSFS :254
QY 642 EAPSVEENITETVAETP-----DFNATADLSALLOPSEVPAVEENAAIVADD--LSAL 694
Db 1255 KPSLSLSPSPSPLEKTPLGERSVNFSLTPNEIKVSAEAEVAPVSEVQVVEEHCAEPE 1314
QY 695 LQAPAEAPAEVENVETVAETSDPHTAADLSALLOPAEV-----PAV-----ENNV 740
Db 1315 DKLEWSPQSQVTSAGHTPYQSPTEKSHL--PTEVIEKPPAVPVSVFSPFSDAKDENE 1373
QY 741 TKTV-----ETPFDNATADLSALLOPSEVPAV--EENAAEITLTPTDSNTSEADALPDF 794
Db 1374 RASVPMDEPVDPD---SESPIEKVLSPLRSPPLIGSESAYESFLSADDKASGRGAESPFE 1430
QY 795 LKQGEETVDKSIYLSEENIPNNAOTSFPSESVGSDAPSEAKYDLAEMYLEIGDR---D 850
Db 1431 EKSQKGGSPDQVSPVSE-----MTSTSLYQDKQEGKSTDPAPIKEDFGQEKKTDD :480
QY 851 AAAETVQKLL-----EEAEGDV 867
Db 1481 VEAMSSQPALADDERKIGDV 1500

RESULT 5

US-08-973-462-22
; Sequence 22, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUIHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 540

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide

US-08-973-462-22

Query Match 6.1%; Score 269; DB 3; Length 540;

Best Local Similarity 20.9%; Pred. No. 2e-11;

Matches 131; Conservative 112; Mismatches 262; Indels 122; Gaps 19;

QY 166 APTAES--QENQNAKALRKTDKDSANAAVKPAYNGKHTTVRKGE-----TVKQIAAAIR 218
Db 18 APSVEEENVEESVAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-- 75
QY 219 PKHITLQO--VADALLKA--NPNVSAHGRLAGSVLHIPLNLRKAEQPKQPTAKPKAETAS 276
Db 76 ----TVEIVAPSVVSVAPSVEEVE-----ENVEESVAENVEESVAENVEESVA 122
QY 277 MPSEPSQATVKEKPEKAAPE-----AKAEKPAVRPEPVPAANTAASETAESAP 331
Db 123 ENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEEVEENVE 182
QY 332 QEAAASALDTPDTGNVSEPVQVSAEEETESGLFGGYSYTLLLAGGGAALIALLLILR 391
Db 183 ESVLENVEESVAENVEESVAENVEES----- 216
QY 392 LAOSKRARTEESVPEE--EPDLDDAADGIBITFAEVETPATPEKPNKNDVNTLALDGE 450
Db 217 ----VAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEE 271
QY 451 SEELSAKQTPDVTETPSNRIDLDFOGLAAQNGILSGALTQDEETQKRADADNNAES 510
Db 272 SVEE-----NVESVAEN-----VEESVAENVEESVAENVEESVAEN 308
QY 511 TDSVVEPETFPNPNP--VEIVDTPESVAOTAENK--PETVDTDFSDNLPNNHIGTBEET 568
Db 309 VEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEEVEENVEESVAENVEESVAENVEES 368
QY 569 ASAKPASPSGLAGFLKASPTILEKTVAEVQTPPELHDFLKVYETDAVAETAPTPDFN 628
Db 369 V-----AENVEEIVAPTVEEIVAP--TVEEIVAPSVVSVAPSVEEVEEN 412
QY 629 AAADLSALLOPAEAPSVVEENITETVAETPDNATADLSALLOPSEVPAVEENAAEIVA 688
Db 413 --VEESVAENVEESVAENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVA 466
QY 689 DDLSSALLOPAEAPAEVENVETVAETSDPHTAADLSALLOPAEVPVEEENVTKTVAEIP 748
Db 467 PSVSVAPSVEEVEENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEII- 520
QY 749 DFNATADLSALLOPSEVPAVEENAAE 775
Db 521 ----VAPSVEESVAPSVEESVAE 539

RESULT 6

US-09-141-047-8
; Sequence 8, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Xu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; FILE REFERENCE: Protein Gene
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141,047A
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 8
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:

Best Local Similarity 22.6%; Pred. No. 7.7e-09;
Matches 184; Conservative 104; Mismatches 273; Indels 254; Gaps 42;

```
QY 40 QIKLIAAASVAAAFQAHAGGNIQSNLDEPFGSITVT-----GEEAKALLGGSV 93
Db 29 RVMLCAMSCLLAGQASA-LGVGDITLHSAHQDADIELLDVGLDADEIEVRLAGADV 87
QY 94 TVSBKGLTAKVHKL-----GDKAVIASSEQAVRDPVLFVFRIGA-----G 133
Db 88 FAA-----AGVERLQFLNELRFSPVLQCRGGNR--IHVSSIRPQVEPVLNLFIVEVRPNG 140
QY 134 AQVREYTAILDPVGSPTKALSQKTHRTAPTAEQENQNAKALRKTDKKUSANAAY 193
Db 141 RIVREFTVLLDPLGTPMLPAARSGIEPQR-----QSSTPAPAPR-----SAAAVV 187
QY 194 KPAY--NGKTHTVRGETVKQIAAIR-PKHLTLEOVADALLKAMP-----NVSAGRLURA 246
Db 188 DPALLEPDEYLRNDNLMAISGLRGAGNADRAQLMEALYQLNPQAFVNAHRH-LRKA 246
QY 247 GSVLHPIPLNRKARQPKPOTAK-----PKAETASMPSEPS-----KQATVE----- 288
Db 247 GVRLRLP--AGYQPERGAPGAVKEAAVEVLPADAADVVENAPALVVEAQRQADAEAEALA 304
QY 289 -----KPVKPKAKVAAPEAK-AEKPAVRP-EPVPAANTAASET 325
Db 305 RQRELSQRMDDLQRLQALQORQDHQVLAELQOQLARRQAVRPAAPPPAA----- 357
QY 326 AASAPQAAASAITDPTDETGNVSEPVQVSEAEETESGLFGGSYTLILLAGGGAALIA 385
Db 358 ---AAP--SVAQPVETPTD-----SQWRMIVLLV--LALLG 389
QY 386 LLLRLAQSKRAKRTESVPEEPDLDADGDIITFAEVETPATPEPAPKNDVNDTL 445
Db 390 VLLL-----RRREEAPVQAVEPKRVA---LNLPLRRAPRPPAAAPAPAK----- 432
QY 446 ALDGESEELSNAQTFDVTETPSNRIDLDFSLAAQNGILSGALTODEETQKRDADW 505
Db 433 -VEQARPPVAA-----PSSPPSPPPAPAPAAAPRAAMAAA----- 466
QY 506 NAIBETSDSYEPETENPNPV-----EVIDTPPEPSVAQTAENKPTVDTDFSDNLPN 560
Db 467 DKLDQAD-IY--IAYRGYQARDLLRQVLAEQPRLS-----ARMKLLVLAEGLD----- 514
QY 561 NHIGTETASAKPSPGLAGLKASSPETILEKTVAEVQTPPEELHDFLKVYETDAVET 620
Db 515 -----AAGF-----DALAETLAGGNSEATDELGRYPAL-LIQMP 549
QY 621 ABETDFNAADLQALLOPAPAPSVENITETVAETDFNATAD-----DLSALL 671
Db 550 ATETPAATTKDDWSD-LPLAEP-----VLQQPDATSGADGFGDLNLDLDWGALE 599
QY 672 QPSEVPVAEENAAEIVADLQALLOPAPAP--AVEENYTE--TVAEISDFHTAADDLSAL 727
Db 600 NPLDNPDLPRRA-----ACKAEPAEELPAPESNLHELDPVAEYB--HLELQO----- 646
QY 728 LQPAEVPVAEENVTKVABIPDNATADLQALQ 762
Db 647 PEPATVPPEASASLDRACIDSGDLQASRIIL 681
```

RESULT 9

US-08-978-277A-4
; Sequence 4, Application US/08978277A
; Patent No. 6582956
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY

```
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,277A  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/665,401  
; FILING DATE: 18-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Richard S  
; REGISTRATION NUMBER: 26,154  
; REFERENCE/DOCKET NUMBER: A30558 - 165/34008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-408-2558  
; TELEFAX: 212-765-2519  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1596 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; US-08-978-277A-4
```

Query Match 5.4%; Score 236; DB 4; Length 1596;

Best Local Similarity 20.2%; Pred. No. 2.6e-08;

Matches 168; Conservative 106; Mismatches 281; Indels 277; Gaps 37;

```
QY 258 IKAEQPKQTAQKAEATAMPSEPSKQATVTK-PVEKPE----- 295
Db 537 IHTSPESADQKQESSASSPEETTCLEKGPLEAPQDGEABEGTSDGKKREGITP 596
QY 296 ---AKVAAPAKAKPAVRPEPVPAANTAASETAAESAPQAAASAITDPTDETGNV 351
Db 597 WASFKMVTPKRVRP-----SESDKEE-ELEKVKSATLSLSDTVSEMQ 641
QY 352 EPVEQVSAEETESGL-----FGGSYTLILAGGGAALIALLLRLAQSKRAKRTESVPE 407
Db 642 DEVKTVGEQKPEEPKRRVDTSVSWEALICVG-----SSKKRAKASSDD 688
QY 408 -----BEPOLD-DAADDGIEITFAE---VETPATPEPAPKNDVNDTLALDGE 450
Db 689 GGPRTLGGDSHRAEASKDKAAGTDAVPASTQEQDQAQSSSPEFA-----GSP 737
QY 451 SE-BELSAKQTFD-VETDTPSNRIDLDFSLAAQNGILSGALTODEETQKRDADWAI 508
Db 738 SEGEGVSTWESFKRLVTPRKKSXKLE---EKADSSVEQLSTEIFPSREES---WVSI 790
QY 509 ES-----TDSVYEPETFNPNPVRVIDTPEPESVAQTAEE---NKPET----- 548
Db 791 KKFIPGRRKKRQADGKQEQATVEDSGPVEINEDDPNVPVAVPLSEVNAVEREKWEAQGNT 850
QY 549 -----VDTDFSDNLPNNHI-----GTEETASAKPASPGLAGFL-----KAS 586
Db 851 LPQLLGAVVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGAM 910
QY 587 SP-ETILEKTVAEVQTP-----EELHDFLKVYETDAVAB--TAPETD----- 626
Db 911 PPVEEVTEKDIIAEETPVLTQTLPGKQADHDMVTSEVDFVSEAVTATETSEALTTEVT 970
QY 627 ----FNAAADDLSALQAPAEAPSVENITET-----VAETPD-----FNATADDLS 668
```

[illegible]

RESULT 10

US-08-635-121-2
; Sequence 2, Application US/08635121

```

; Patent No. 5910442
;
; GENERAL INFORMATION:
;
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 2

```

; ;
; ;
; ;
; ;
; ;

CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York

Query Match 5.3%; Score 235; DB 2; Length 1346;
Best Local Similarity 20.2%; Pred. No. 2.4e-08;

Matches 171; Conservative 102; Mismatches 268; Indels 304; Gaps 37;	
QY	258 IKAEQPKPQTAKPAETASMPSEPSKQATVEK-PVEKPEAKVAAPAEKAEKPAVRPP-- 314
Db	151 IHTSPESADQKQESSASPEEPETTCLEKGLEAPRM-----GKLRKELLRGEKKR 204
QY	315 -----VPAANTAASETA---AESAPQ---FAASAIDTPTDTGNVAVSEPFVEQVSAABE 362
Db	205 KDHSLGILQKDGTQETVRRPSPDSKBEELKVKKSATLSSTSDTSVSMQDEVKTVGEEQK 264
QY	363 TE-----SGLFGGSYTLILLAGGAALIALLLLRLA 393
Db	265 PEEPKRRVDTSVSWEALICVGSSKKRKAKSSDIRPRTL---GGGQ-----S 310
QY	394 QSK----RAARTESVPEEPDLDDAADGGIEITFAEVETPATPEPA----- 436
Db	311 QSRGGQQRQSRRTDAVASTQEQQA-----QSSSPPEPAGSPSEGEVSTWE 358
QY	437 -----PKNDVNTTLADGESEELSAAKQTFDVEDTP--SNRID--LDFDLSLAAQAQG 485
Db	359 SFKRLVTPRKXKSKL-----EKEAGRTLTVVGACPLRSNRKVEKNLGFPLRNSPDG 411
QY	486 ILSGALTQDEBTQKGAADWNATTESTVSVEPETFNPYNPVEIVIDTPEPESVAQTAEN- 544
Db	412 GRKGQMRQEQA-----TVEDSGFEINDEDFVPAVVPVLSYED 450
QY	545 --KPEITVDTFDSNLPs-----NNHI-----GTEETASAKPASPSGLAGF 582
Db	451 AVEREKMEAQNAELPSCWGCVSSELSKTLVHTVSVAVIDGTRAVTSEERSEPSWISAS 510
QY	583 L-----KASSP-ETILEKTVAEVQTP-----BELHDFLKVEYETDAVE--TA 621
Db	511 VTEPLEHTAGEAMPVVEVTEKDIIAETPVLTQTLPEGKDAHDDMTVSEVDTTSEAVTA 570
QY	622 PETPD-----FNAADDLSSALQAPASPSVEENITET-----VAETPD----- 659
Db	571 TETSEALRTBEVTEASGAEEETDMVSAVSQUTDSPDTTEATPVQVEGGVLDTTEBERQ 630
QY	660 -----FNATADLSALLQPSRVA-----VEENA----- 683
Db	631 TQAILQAVQKVK---EESQVPATQTVORTGSKALEKVEVEEEDSEVLASEKDVMPKG 687
QY	684 -----AEIVAD-----DLSALLOPAEAPAVEENV 707
Db	688 PVQEGAGAEHLAQSGETQATPESILEVPEVTDVDVHVATCQVIKLQQLMEQAVAPESSETL 747
QY	708 TETVAETSDPHTAADLSALLQAPAEVPAVEENVTKTVAE---IPDNATADLSALLQPS 764
Db	748 TDSINGSTPLADSDTADGTQQDETDSQDSKATAAVRQSVQVTEEEAATAQKEPSTLPN 807
QY	765 EVPAVENAAEITLET--PDSNTSEADALPDFLKD--GREETVDWSTYLSEENI----- 814
Db	808 NVPAQEEHGEPRGRVLEPTQQBELAAAVPVQKTVGQEGEVDW---LDGEKVKGEQEV 864
QY	815 ----PNN----ADTSPFSEVSG--SDAPSEAKYDLAEWYLEIGRDAAAE-----TWQKL 859
Db	865 FVHSGPNSQKAADVTDSEVMGVAGCQEKESTEVQSLSEEGEMETDVEKEKETPEQV 924
QY	860 LEEAE 864
Db	925 SERGE 929

RESULT 11
US-08-978-277A-2
; Sequence 2, Application US/08978277A
; Patent No. 6582956
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue &

STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASCSQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,277A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/665,401
FILING DATE: 18-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Richard S
REGISTRATION NUMBER: 26,154
REFERENCE/DOCKET NUMBER: A30558 - 165/34008
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2558
TELEFAX: 212-765-2519
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-978-277A-2

Query Match 5.3%; Score 235; DB 4; Length 1346;
Best Local Similarity 20.2%; Pred. No. 2.4e-08;
Matches 171; Conservative 102; Mismatches 268; Indels 304; Gaps 37;
QY 258 IKAEQPKQAKPAETAMPSEKQATVEK-PVEKPAKVAEPAKAEKPAVPEP-- 314
DB 151 IHTESPESADEQKGESSAPSEPEPTCLEKPLEAPRM-----GKLKELLRGEK 204
QY 315 -----VPAANTAASETA---AESAPQ---EAAASAIPTDPTDGTNAVSEPEVQSAEE 362
DB 205 KDHSIGILQKQDGTQETVRRPSESKEELEKVKKSATLSSTSTVSEMDEKVTGVEOK 264
QY 363 TE-----SGLFQGSYTLILLAGGGAALIALLLRLA 393
DB 265 PREPKRVDTSVSWAELICVGSKKPKARKASSSDIRGPTL---GGGQ-----S 310
QY 394 QSK---RARRTEHSVPEEPDLDAAADGIEITFAVETPATPEPA----- 436
DB 311 QSRGQQRQRRTDAVPASTQEQDQ-----QGSSEPEPAGSPSEGEVSTWE 358
QY 437 -----PKNDVNDTLALDGESEELSAKQTFDVEDTDP--SNRID--LDQSLAAQNG 485
DB 359 SFKRLVTPKSKSL-----EKEAGRTLVVGAGCPRLNRVKNLGFPLRNSPDG 411
QY 486 ILSGALTODEETQKRADADWNAIESTDSVYEPETFPNYPNVEIVDTPEPESVQAEN- 544
DB 412 GRKGQNGRQEQ-----TVEDSGPVEINDEPDPVAVPLSEYD 450
QY 545 --KPEVTDFTDSDNLPS-----NNH-----GTEETASAKPASPGLAGF 582
DB 451 AVEREKWEAQGNALPSCWGCVVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWIS 510
QY 583 L-----KASSP--ETILEKTVAEVQTP-----EELHDFLKVYETDAVAE--TA 621
DB 511 VTEPLEHTAGEAMPPEVEVTEKDI IAEETPVLQTLPKGDADDDMTSEVDFTEAVTA 570

QY 622 PETPD-----FNAAADDLSALLQPAEAPSVEENITET-----VAETPD----- 659
DB 571 TETSEALRTEBVTASGAETTDWVSAVSQLTDSPTTTEATPVQVEGGVLDTEEBRQ 630
QY 660 ----FNATADDLSALLQPSVPA-----VEENA----- 683
DB 631 TQAILQAVADKVK---BESQVPATQTVORTQSKALEKVEEVEEDSEVLASEKEDVMPKG 687
QY 684 ----AEIVAD-----DLSALLQPAEAPAEVENV 707
DB 688 PVQAGAGHLAQGSETQCATPESLEVETADVHVATCVIKLQQLMEQAVAPESSETL 747
QY 708 TETVAETSDPHTAADDLSALLQPAEVPVENVTKTVAE---IDFNATADDLSALLQPS 764
DB 748 TDSETNGSTPLADSDTADGTQDETIDSDSKATAAVRQSQVTEEEAATAQKEEPTILPN 807
QY 765 EYPAVEENAAEITLET--PDSNTSEADALPDFLD--GEETVDWMSIVLSENI----- 814
DB 808 NYPAGEHGEHGFGRDVLLEPTQOELAAAVPVQKTEVQGEVDM---LDGEKVKEQEV 864
QY 815 ----PNN---ADTSFPSESVG--SDAPSEAKYDLAEMYLEIGDRDAAAE-----TVQKL 859
DB 865 FVHSGPNSQKAADVTYDSEVGVAGCQEKESTEVQSLSEBSEMETDVEKEKRETKPEQV 924
QY 860 LEEAR 864
DB 925 SEEGE 929

RESULT 12
US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 5.3%; Score 235; DB 4; Length 2137;
Best Local Similarity 17.1%; Pred. No. 4.8e-08;
Matches 154; Conservative 177; Mismatches 407; Indels 160; Gaps 26;
QY 11 PMTKPTDCTRSNRITQPPHFG-----YILKNRQIKLIAASVAVASQ 55
DB 750 PQGLKF-DASTNSIVGTPTQIGTNTITESTDASGNKTTTINYEVRNSASDSTSTS-- 806
QY 56 AHAGLGLNIQNLDEPFGSITVTEGEEAKALLGGSVTVSEKGLTAKVHKLGDRAVIAV 115
DB 807 ----INVSVTSISNSTLS-----DSVKASOSLSTSK--SISELSAST 845
QY 116 SSEQAVRDPVLVFRICAGAGQVREYTAILDPVGYSPKTSALSDGKTHRTAPTASQENQ 175
DB 846 SNSTSI-----QASESASTSKQLESASTSTSDSASARKSESTSKS----- 888
QY 176 NAKALRKTDDKDSANAAVKPATNGKHTVTKGETVKQIAAARPKHLTLEQVAD-ALLKA 234
DB 889 -TSLSESTSTSVSDSASVSTSEASTSTSVSGSTSTSTSDS-----TSTSTSDSASIK 941
QY 235 NPNVSAHGLRAGSVLHINLRKAEQPKQATKPAETASMPSEPPSKQATVEKPEVKP 294

Patent No. 6245337
 GENERAL INFORMATION:
 APPLICANT: St. Gene III, Joseph W.
 APPLICANT: Falkow, Stanley
 TITLE OF INVENTION: Haemophilus Adherence and Penetration
 TITLE OF INVENTION: Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 FILING DATE: 25-AUG-1994
 APPLICATION NUMBER: US/08/296,791
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Trecartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1848 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 US-08-296-791-6

Query Match 5.3%; Score 232.5; DB 3; Length 1848;
 Best Local Similarity 19.2%; Pred. No. 5.9e-08;
 Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;
 QY 35 LKNNRQIKLIAASVAVASPOAHG-----LGGINI-QSNLDEPPFGSITVT 80
 DB 766 LYSGRNVANITSNITASNAAQVHIGYKGTGTVCVRSYDTGYVTCHNSNLSEKALNSFNPT 825
 QY 81 GEEAKA-LLGGGVTVSVSEKGLTAKVKLGDKAVIAVSSEQAVRDPVLFRIGAGAAQVREY 139
 DB 826 NLRGVNUTENASFTLGRANLFGTIQSIGTQVNLKNSH-----WHLTGNSNVNQL 877
 QY 140 TAILDPVGYSPKTSALSDGKTHRTKAPTAESENQNAKALRK-----TDKK 186
 DB 878 N-----LTNGHILHNAQNDANKVTYNTLTVNSLSGNGSFYVWVDFTNK 922
 QY 187 DSANAAVPAVNGKHTV--RKEITVKGIAAIPKH--LILEQVADA-----LLKANP 236
 DB 923 SNKVVWNSKATGNFTLQVADKTGE-----PNHNETLFDASNATRNLEVTLANG 972
 QY 237 NV-----SAHGRU-----RAGSVLHILPNLRKIAEQPKPTAK-----PKA 272
 DB 973 SVDRGAWKYKLRVNGRDLVNPVEKRNQVDTNITPNDIQADAPSAQSNNEEIAV 1032
 QY 273 ETASMPSPSKQATV--EKPVKPEAKVAAPAEKAPVAPPEVPAANTAA-----322
 DB 1033 ETPVPPPPAPATSAIAEQPETRP--AETAQPAEETNTANGSTETAPKSDTATQENPNS 1091
 QY 323 ---SET-----AASNPQAAASADTPTDETGNVASEVPEQVS 358
 DB 1092 SVPSETTEKVAENPPQENETVAKNQEATEPTPQNGEVAKEDQPTVEANTQNETQSEG 1151
 QY 359 ABEETESGLFGSYTLLLAGGAALIALILLRLAQSKRRARTEESVPEEPD--LDDAA 416
 DB 1152 KTEETQT-----AETKSEPTSVTSVENQPEKTSQST 1184

QY 417 DDGIEI---TFARVET-----PATPEAPKNDVNDTLALDGSSEBELSAKP 459
 DB 1185 EDKVVVEKEKAKVETEETOKAPQVTSKEPKQAPAEPEEPTDTNAAEQALQQTPTT 1244
 QY 460 TFDVETDTPSNRDLDFSLAAQNG-----ILSGALTQDEETQKADADNNAIESIDS 513
 DB 1245 VAAAEITSPNSKPAEETQOPSEKTNABFVTPVVSSENTATQTEETEETAKVE---KEKTQE 1301
 QY 514 VYBPETENPNVPIVIDTPEPEVAQTAENKPTVDTDFSDNLPNSNNHIGTEETASAKP 573
 DB 1302 V--PQVASQSPKQ---EQPAKPAQAT---KPOAEP--RENVLTKVVG-EQPOQAQP 1350
 QY 574 ASPS---GLAGFLKASSPETILEKTVAEVQPEBELHDFLKYVETDAVAETAPETPDFNAA 630
 DB 1351 QTQSTAVPTTGETAANSKPAKPAQAKPQTPAPARENVTVTNKEPQSTSATVSTEQPA 1410
 QY 631 ADDLSALLQPAEAPSVENITETVAETPDFNATADDLSALLQPEVPAVENAAEIVADD 690
 DB 1411 KETSSNVEQPAPENSINTGSATTMTET-----AEKSKPKQME---TVTEND 1453
 QY 691 LSALLQPAEAPAVENVTETVAETSDFTHTAADDLSALLQPAEVPVAVENVTKTVAEIPDF 750
 DB 1454 ---RQP-EANTVADNSVANNSESE--SKSRRRSYSQPKETSAEETTVASTQ-----1500
 QY 751 NATADDLSALLQPESE---VPAVEENAAEITLTETPDSTNTSEADALPDFLXGEEETVDMWI 807
 DB 1501 ETTVDNSVSTPKPRSRRTSRSVQTSNVE--PVELPTENAENAVNQ-----1544
 QY 808 YLSEENTPNAD-----TSFSPSEVSGDAPSEAKYDLAEMYLEIGDRDAAAEVOKLLEE 862
 DB 1545 --SGNNVANSQPALRNLTKNTNAVISNAMAQAF---VALNVG---KAVSQHISQLEMMN 1596
 QY 863 AEG 865
 DB 1597 NEG 1599

RESULT 15
 US-09-839-996-6
 ; Sequence 6, Application US/09839996
 ; Patent No. 6642371
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Gene III, Joseph W.
 ; Falkow, Stanley
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration
 ; Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/839,996
 ; FILING DATE: 20-Apr-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791
 ; FILING DATE: 25-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1848 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6

Query Match 5.3%; Score 232.5; DB 4; Length 1848;

Best Local Similarity 19.2%; Pred. No. 5.9e-08;

Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

```
QY 35 LKNROIKLIAASVAVAAFPQAHAG-----LGGLNI-QSNLDEPFGSITVT 80
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 766 LYSGRVANITSNITASNNAQVHIGYKTGTVCRSDYTGTVYVICHNSNLSEKALNSFNPT 825
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 81 GBEAKA-LLGGGSVTVSEKGLTAKVHKLGDKAVIAVSEQAQRDPVLVFRIGAGAVREY 139
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 826 NLRGNVNLTENASFTLGKANLFGTIQSIGTSQVNLKENS-----WHLTGNSNVNQL 877
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 140 TAILDPVGYSPKTSALSCKTHKTAFTAESQENQAKALRK-----TDKK 186
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 878 N-----LTNGHIHLNAQNDANKVTTYNTLTVNSLSGNSGFYVWVDFTNK 922
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 187 DSANAARKPAYNGKTHV--RKGETVKQIAAAIRPKH--LTLQVADA-----LLKANP 236
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 923 SNKVVNKSATGNTLQVADKGE-----PNNELTLFDASNATRNNLVLIANG 972
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 237 NV-----SAHGR-----RAGSVLHPIPNLRKAEQPKQTAK---PKA 272
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 973 SVDRGAWKYLRNVNGRYDLNPEVEKRNQTVDTNITTPNDIQADAPSAQSNNEEIARV 1032
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 273 ETASMPSEPSKQATV--EKVEPEKAAVAAPEAKPAVRPEVPAANTAA-----322
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1033 ETPVPPAPATESAIASEQPETRP-AETAQAPAMEETNTANSTETAPKSDTATQENPNS 1091
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 323 ---SET-----AESAPQEAASAIPTDPTDETGNVSEPEVQVS 358
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1092 SVFSETEKVAENPPPOENETVAKNEQATEPTQNGEVAKEQDPTVEANTQTNEATQSEG 1151
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 359 ABEETESGLFGGSYILLAGGGAALLALLLLRLAQSRRRTESVPEEEDP--LDDAA 416
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1152 KTEETQT-----AETKSEPTSVTVSENOPEKTVSQST 1184
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 417 DQGIET---TFARVET-----PATPEPAPKVDVNDTLALDGESEELSAKQ 459
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1185 EDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAPAEPEVPTDTNAEEAALQQTQPTT 1244
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 460 TFDVETDTPSNRIDLDFDSLAAQNG-----ILSGALTQDEETOKRADADWNAIESTDS 513
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1245 VAAAEITSPNSKPAETQOPSEKTAEPVTPVVSSENTATQPTETEETAKVE---KEKTQE 1301
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 514 VPEPTFPNPNVEIIVDTPESVQAETAKNKETVDTDFDNLPSNNHIGTEETASAKP 573
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1302 V--PQVASQESPKQ---EQPAKPAQCT--KQAPAEPA--RENVLTJKNVG-EPQPAQAP 1350
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 574 ASDPS---GLAGFLKASPETILKTVAEVQTPPELHDFLKVETDAVAETAPETPDFNAA 630
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1351 QTQSTAVPTTGETAANSKPAKPAQAKPQATEPARENVTNTKPEQSQTSATVSTEQPA 1410
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 631 ADDLSALLOPAEAPSVVEENITETVAETPDFNATADDLSALLOPSEVPAPVEENAAEIVADD 690
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1411 KETSSNVEQAPENSINTGSAITWTET-----AEKSDKQME---TVTEND 1453
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 691 LSALLQPAEAPAVEENVETVAETSDPHTAADDLSALLQPAEAPAVEENVTKTVAEIPDF 750
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1454 ----RQP-EANTVADNSVANNSESSE--SKSRRRRSVSPKETSABEITTVASTQ-----1500
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 751 NATADDLSALLOPSE---VPAVEENAAEITLETDPDNTSEADALPDFLKDGBEETVDSI 807
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1501 ETTVDNSVSTPKPRSRRTSRVQTSNYSY-PVELPTENAENVQ-----1544
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
QY 808 YLSEENIPNNAD-----TSFPSESVDGSDAPSEAKYDLAEMYLEIGDRDAAAETVOKLEE 862
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1545 --SGNNVANSQPALRNLTSKNTNAVVISNMAKAQF-----VALNVG--KAVSQHISQLEMN 1596
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 863 AEG 865
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1597 NEG 1599
```

Search completed: September 2, 2004, 19:21:56
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:20:52 ; Search time 144 Seconds
(without alignments)
1925.398 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRCPMTMTKTDTCT.....EEAEGDVLKRAQLAQLGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3434	78.0	801	12	US-10-282-122A-65373, A
2	2391.5	54.3	503	12	US-10-282-122A-65571, A
3	381.5	8.7	919	12	US-10-282-122A-66451, A
4	358	8.1	1095	14	US-10-128-714-8305, A
5	315.5	7.2	1786	9	US-09-742-096-3, A
6	315.5	7.2	1787	12	US-10-415-253-2, A
7	315	7.2	1000	14	US-10-128-714-3305, A
8	276.5	6.3	1772	16	US-10-437-963-143280, A
9	276.5	6.3	1844	16	US-10-437-963-143282, A
10	269	6.1	540	9	US-09-742-096-22, A
11	269	6.1	2364	12	US-10-205-331-66, A
12	260.5	5.9	1444	16	US-10-437-963-150334, A
13	258	5.9	1859	15	US-10-369-493-19525, A
14	253.5	5.8	3507	15	US-10-369-493-5784, A
15	253	5.7	1367	9	US-09-801-368-108, A

16	251	5.7	1348	12	US-10-282-122A-56877, A
17	250.5	5.7	630	9	US-09-742-096-5, A
18	244	5.5	1616	10	US-09-820-843A-16, A
19	244	5.5	1616	12	US-10-282-122A-63593, A
20	242.5	5.5	530	16	US-10-437-963-117065, A
21	238	5.4	2665	9	US-09-864-761-34248, A
22	238	5.4	3664	12	US-10-263-929-143, A
23	238	5.4	3664	14	US-10-177-293-423, A
24	238	5.4	3664	16	US-10-408-765A-2287, A
25	236	5.4	1596	9	US-09-902-432-4, A
26	236	5.4	2478	9	US-09-815-242-5816, A
27	236	5.4	2478	9	US-09-815-242-12967, A
28	236	5.4	6642	15	US-10-369-493-5013, A
29	235	5.3	1346	9	US-09-902-432-2, A
30	233	5.3	740	15	US-10-369-493-3700, A
31	233	5.3	1665	12	US-10-282-122A-71690, A
32	232.5	5.3	1848	10	US-09-839-996-6, A
33	232.5	5.3	1848	12	US-10-645-655-6, A
34	232.5	5.3	1848	14	US-10-080-505-6, A
35	232.5	5.3	1848	16	US-10-687-046-6, A
36	230	5.2	2435	12	US-10-282-122A-47453, A
37	229.5	5.2	2481	12	US-10-282-122A-43762, A
38	228.5	5.2	2368	9	US-09-815-242-5635, A
39	228.5	5.2	2368	9	US-09-815-242-12389, A
40	227.5	5.2	1831	12	US-10-282-122A-71033, A
41	227.5	5.2	2472	9	US-09-815-242-5064, A
42	226.5	5.1	1178	14	US-10-128-714-8240, A
43	226.5	5.1	1879	9	US-09-971-536-70, A
44	226.5	5.1	2402	16	US-10-661-809-20, A
45	226	5.1	570	12	US-10-415-253-4, A

ALIGNMENTS

RESULT 1

US-10-282-122A-65373
; Sequence 65373, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

Db 834 -----NLDPEKATPSPSARDAAVASALDGD-----ADDDPDF 866

Qy 825 ESVGSDAPSEAKYDLAEMYLEIGDRDRAAAETVQKLLAEAGDVLKRAQA 873

Db 867 LSGADEAAT--KLDLARAYIDMGSEGDILDEVL--AEGNDSQAE 911

RESULT 4

US-10-128-714-8305

; Sequence 8305, Application US/10128714

; Publication No. US20030119013A1

GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wendi

; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos

; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lenieux, Sebastien M

; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and

; TITLE OF INVENTION: Methods of Use

; FILE REFERENCE: 10182-018-999

; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8305

; LENGTH: 1095

; TYPE: PRT

; ORGANISM: *Aspergillus fumigatus*

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (1095)..(1095)

; OTHER INFORMATION: X= any amino acid

US-10-128-714-8305

Query Match 8.1%; Score 358; DB 14; Length 1095;

Best Local Similarity 22.1%; Pred. No. 1e-13;

Matches 233; Conservative 140; Mismatches 412; Indels 270; Gaps 44;

Qy 38 NRQIKLIAASVAVASFOAHAGLGLNQTOSMLDEPFGSGITVTGEEAKALGGGSGVTVSE 97

Db 91 NNILTVGSSSIAAAPGAHA-----VEKETETPLENGASEKVEETAKEPGVQSTTTTTE 144

Qy 98 KGLTAKVHKLGDKAVIA-----VSSEQAVRD--PVLVFRIGAGAVRE--YTAILDVPGVS 149

Db 145 AVKDEVEKTDSDVVVEKDLHVESEPATQEVKPKLVSEPSKPSDIHEDAEIQTASNAAE 204

Qy 150 PKTKSALS-----DGKTHRTKTAFTAESQENQAKARCTDKOSANAAVKPAYNGK 200

Db 205 PKTNGTHSSEQISGSPQDDAVTEKVEPVEEKAADLTTKTKVANKPBEIPSTLSAEKPE 264

Qy 201 THTVRKGE-----TVKQIAAIRPKHL-----TLEQVADALLKANPNVSAHGLRL 245

Db 265 TKVEKEADQKLDLEIVPAITVEKTEAKESPESTLEKSQBEAAPAQ--TVTVEEPLV 323

Qy 246 AGSVLHLPNLRIRKAEQ--PKPOTA--KPKAETAS---MPSEP---SKQATV----- 287

Db 324 DSKPVEISSAESTKAEKAPTMTTDEAPLAKSEKTVPAEPIAEKVKQAIIGDEPLSRN 383

Qy 288 ---EKVKEKP--EAKVAPEAKAEKPAVRPEPVAANTAASETAESAPEQEAASAI--DTP 342

Db 384 NVEEQPKQKSVKAEAKADEAVLEEP-----IKESAPEEIPEDSRNAVADAP 430

Qy 343 TDETG-----NAVSEPVQVSAEEETESGLFGGSYTLLLAGGGAALIALLLRL 392

Db 431 VTSESTTEKVEVTAPDAEKEPALKESAKEPIQEIBPEASEKEVAE-----TPV 479

Qy 393 AQSKRAARTESV--PREEPLDDAADDGI-----EITFA-EVETPATPEPAP 437

Db 480 TESSTTEKVEVTALDAEKPALEESTKEPIHEEQTIETVTVTAETAAKEPAQATEEPPVA 539

Qy 438 KNDVNDTILADGSESE-----ELSAKQTFDVT-----DTFSNRKIDLDLDFDSLAAAQNGI 486

Db 540 NESINK--EVEATTEETAGQESAKEPISNETAIRVTOADESTKELSGTF---AAAEIT 594

Qy 487 LSGALTODEET-----OKRADADWNAIESTD----- 512

Db 595 GKGAETHTEVSTAAGPAKEFVNDASAVKSAPTEVTQETPKDDSAQVVVVGKEABEQIAER 654

Qy 513 -SVYEPETFPNPNVEIVIDTPEP-----ESVAQTAENKPETVTD 550

Db 655 AAETESAANPEATEETSTTAERPAKEPGESEETPCQILPAAEPKSEKATAEBAKEESTIE 714

Qy 551 TDFSDNL-----PSNNHIGTEETASAKPASPSGLAGFLKASSPETTILEKTVAEVQTPEE 604

Db 715 T-VSEPLEADVKEPAQDDPATETLTEDKPVTOESVAE--AAPKPEVSEPVVSEGTAKTS 771

Qy 605 LHDFLKVYETDVA-----ETAPETPDFNAAD--DLSALLOPAEA 643

Db 772 AVDEAESSEKPAVVTETKEPGFEBHPVTRELAKSESEASDKAAKSGEAPAAPETILEA 831

Qy 644 PSVEENITETVAETPDFNATADDLSALLQSPSEVPAVEENAEIV-----ADDLSALLQOP 697

Db 832 PAPETVPPEVSPSEFTEAVAKE-----PATVESTQETPTTGTGEQAKGEVAINEP 881

Qy 698 AEAPAVEENVVTETVAETSDFTAAADLSALLQPAEVPVEENVTKTVAEIPDFNATADDL 757

Db 882 TEATATEEAVPEEVSEVKG--EIEEPVAAVKKSELPADEPTLCNDGCS--PEEITTEEA 936

Qy 758 SALLQSPSEVPAVEENAEIITLETDPDSNTSEADALP-----DFLKDGE-----EETVDWS 806

Db 937 VQKTKASDEPVEETNATEVTIKEPTA--TETTESAPVKEATETVKEPEAVAPVABEPVKTA 995

Qy 807 I-----YLSEENI-----PNNADTSFPSESVGSDAPSEAKYDLA-----EMYL 844

Db 996 IAEQDNTTEKLAETVSKDVTEPSEVPAEKVVPSEPEATEEPVSEPTSEATERAEPEASI 1055

Qy 845 EIGDRDAAAE--TVQKLLAEAGDVLKRAQALAE 877

Db 1056 QVPAQDEVADVAAIEEQEQAEEVPEVTKPAAEPAAE 1090

RESULT 5

US-09-742-096-3

; Sequence 3, Application US/09742096

; Patent No. US20020155441A1

GENERAL INFORMATION:

; APPLICANT: DRUILHE, PIERRE

; APPLICANT: DAUBERES, PIERRE

; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

; FILE REFERENCE: 200773USODIV

; CURRENT APPLICATION NUMBER: US/09/742,096

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: US 08/973,642

; PRIOR FILING DATE: 1998-02-06

; PRIOR APPLICATION NUMBER: PCT/FR96/00894

; PRIOR FILING DATE: 1996-06-12

; PRIOR APPLICATION NUMBER: FR 95/07007

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3

; LENGTH: 1786

; TYPE: PRT

; ORGANISM: *P. falciparum*

US-09-742-096-3

```
Query Match          7.2%; Score 315.5; DB 9; Length 1786;
Best Local Similarity 20.2%; Pred. No. 9.5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSNLDPEPFGSITVTGEEKALLOGGSVTVSEKGLTAKVHKLGDKAV---IAVSEQA 120
Db 242 NVEENENDGSGVASSVEES-----IASSVDESIDSSTEENVAPTVEEI 286
QY 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDBGKTHKRTAPTAE--QENQNAK 178
Db 287 VAPSVV-----ESVAPSVSEESVEENVEES 310
QY 179 ALRKTDKDSANAAPKPAVNGKTHTVRKGE-----TVKQIAAAIRPKHLTLEQ-VADALL 232
Db 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRLRAGSVLHPLNLRKIKAPQPKQAKPAETASMPSEPSKQATVEKPV 291
Db 365 ESVAPSVSEESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPPAANTAASETAESAPQEAASAITDPTDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESVEENVEESVAENVEESVAENV 475
QY 347 GNAVSEPFVEQVSAEERTESGLFGSYTLLLAGGGAALIALLLLRQAQKRARRTEESVP 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADDGIITFAEVETPATPPAPKNDVNDTLALDGESEELSAKQTFDET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGILSGALTQDEETQKRADADWNALESSTDVYEPETFPNYP 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPEPESVAQTAENK-PETVDTDFDNLPSNNHIGTETASAKPASPGLAGFL 583
Db 602 TVEEIVAPSVVESVAPSVSEESVEENVEESVAENVEESVAENVEESV-----VAPSVEE 647
QY 584 KASSPTILEKTVAEVQTPPELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVSEESVEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVAPTDFNATADDLSALLQPSVPVPAVEENAAEIVADDLSALLQPAEPAPV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESV 759
QY 704 EENVETVAETSDFHFTAADDLSALLQPAEPVPAVEENVTKTVAEIPDFNATADDLSALLQ 763
Db 760 EENVEESVAENVB-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSVEE 805
QY 764 SEVPAVEENAAEITLETPTSNTSEADALPDFLKDGEETVDMWSIYISEENIPNNAUTSFP 823
Db 806 SVAPSVSEESVA-----ENVATNLS-D-NLLSNLGGIETEIEIKDSILNEIBEVENVTITI- 859
QY 824 SESVGSDAPESEAKYDLAEMYLEIGDRDAAAEIVQKLEEAEGDVLKRA 871
Db 860 LENV-BETTAESVTTFSNILEEIQENTITNDTIEEKLLEHENVLSAA 906
```

RESULT 6

```
US-10-415-253-2
; Sequence 2, Application US/10415253
; Publication No. US20040067236A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Joe
; APPLICANT: Druilhe, Pierre
; TITLE OF INVENTION: Immunogenic Compositions Comprising
; TITLE OF INVENTION: Liver Stage Malarial Antigens
; FILE REFERENCE: B45250
; CURRENT APPLICATION NUMBER: US/10/415,253
```

```
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1787
; TYPE: PRT
; ORGANISM: K1 Parasite Clone
; US-10-415-253-2

Query Match          7.2%; Score 315.5; DB 12; Length 1787;
Best Local Similarity 20.2%; Pred. No. 9.5e-11;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSNLDPEPFGSITVTGEEKALLOGGSVTVSEKGLTAKVHKLGDKAV---IAVSEQA 120
Db 242 NVEENENDGSGVASSVEES-----IASSVDESIDSSTEENVAPTVEEI 286
QY 121 VRDPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDBGKTHKRTAPTAE--QENQNAK 178
Db 287 VAPSVV-----ESVAPSVSEESVEENVEES 310
QY 179 ALRKTDKDSANAAPKPAVNGKTHTVRKGE-----TVKQIAAAIRPKHLTLEQ-VADALL 232
Db 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRLRAGSVLHPLNLRKIKAPQPKQAKPAETASMPSEPSKQATVEKPV 291
Db 365 ESVAPSVSEESVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPPAANTAASETAESAPQEAASAITDPTDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESVEENVEESVAENVEESVAENV 475
QY 347 GNAVSEPFVEQVSAEERTESGLFGSYTLLLAGGGAALIALLLLRQAQKRARRTEESVP 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADDGIITFAEVETPATPPAPKNDVNDTLALDGESEELSAKQTFDET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGILSGALTQDEETQKRADADWNALESSTDVYEPETFPNYP 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPEPESVAQTAENK-PETVDTDFDNLPSNNHIGTETASAKPASPGLAGFL 583
Db 602 TVEEIVAPSVVESVAPSVSEESVEENVEESVAENVEESVAENVEESV-----VAPSVEE 647
QY 584 KASSPTILEKTVAEVQTPPELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVESVAPSVSEESVEEN-VEESVAENVEESVA 704
QY 644 PSVEENITETVAPTDFNATADDLSALLQPSVPVPAVEENAAEIVADDLSALLQPAEPAPV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVSEESV 759
QY 704 EENVETVAETSDFHFTAADDLSALLQPAEPVPAVEENVTKTVAEIPDFNATADDLSALLQ 763
Db 760 EENVEESVAENVB-----ESVAENVEESVAENVEESVAPTVEEI-----VAPSVEE 805
QY 764 SEVPAVEENAAEITLETPTSNTSEADALPDFLKDGEETVDMWSIYISEENIPNNAUTSFP 823
Db 806 SVAPSVSEESVA-----ENVATNLS-D-NLLSNLGGIETEIEIKDSILNEIBEVENVTITI- 859
QY 824 SESVGSDAPESEAKYDLAEMYLEIGDRDAAAEIVQKLEEAEGDVLKRA 871
Db 860 LENV-BETTAESVTTFSNILEEIQENTITNDTIEEKLLEHENVLSAA 906
```

RESULT 8

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A
APPLICANT: Barbazuk, Brad

```

1  APPLICANT:  Li, Ping
2  TITLE OF INVENTION:  Rice Nucleic Acid Molecules and Other Molecules Associated With
3  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
4  FILE REFERENCE:  38-21(53221)B
5  CURRENT APPLICATION NUMBER:  US/10/437,963
6  CURRENT FILING DATE:  2003-05-14
7  NUMBER OF SEQ ID NOS:  204966
8  SEQ ID NO 143280
9  LENGTH:  1772
10 TYPE:  PRT
11 ORGANISM:  Oryza sativa
12 FEATURE:
13 OTHER INFORMATION:  Clone ID:  PAT_MRT4530_44203C.1.pep
14 US-10-437-963-143280

```

Query Match	6.3%;	Score 276.5;	DB 16;	Length 1772;
Best Local Similarity	21.2%;	Pred. No. 2.7e-08;		
Matches 212;	Conservative 142;	Mismatches 379;	Indels 267;	Gaps 42;
Qy	6	LPRCPMWTKFTDCTGRNRIQPPHTRGYILKNRQ----	IKLIAASVAVAASFOAHAGL	60
Db	481	VPQEHSGITEKPSQTSADDVKPSESTAAVYKQWQNRNIGFKQPOSSVTDAKTPESR-GM	539	
Qy	61	GGLNIQNLDREPPSGSITVTGERAKALLGGSVTVSEKGLTAKVHKLGDKXAVIA-----	114	
Db	540	GKRLWQRNVP-----TEKQSQS-----GASVVTSPQSVASGNTPLQSSVAVYKKPWQRS	587	
Qy	115	-----VSSEQAVRDPVLVFRIGAGAQVREYTAILDVPGVSPKTKSALSDGKTHRK	164	
Db	588	VSREKEPEKDISNNKPLQNRVL-----AEEAEKTNV-----TADNKSQIIQDKKNG	634	
Qy	165	TAPTAESQENQNAKLRKTKDKSANAAV-----	KPAYNGKTHT	203
Db	635	ATTENSQISESAKKLPLTANKQADTSTKELSESILAVASSQITEPPSKPSN-----	689	
Qy	204	VRKETVYQTAATAIRPKHLTEQVADALLKANPNVSAHGRAGSVLHILNLRIRKAEQP	263	
Db	690	--TAENEKPSQTDIATDELPLTQSGVEAMSPSPSDAAHOISEREIL-----TDKLP	739	

Db 1310 LDPDGMF-----RGVSAPVVFHVAQSRRELVWK--QRLAQ 1342

RESULT 10

US-09-742-096-22
; Sequence 22, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773US01V
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-09-742-096-22

Query Match 6.1%; Score 269; DB 9; Length 540;
Best Local Similarity 20.9%; Pred. No. 1.5e-08;
Matches 131; Conservative 112; Mismatches 262; Indels 122; Gaps 19;
Qy 166 APTAES--OENONAKALRKTDKDSANAAVDPYNGKTHTVRKGE-----TVKQIAAAIR 218
Db 18 APSVEESVEENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-- 75
Qy 219 PKHITLQ--VADALLKA--NPNSAHGRRLRAGSVLHPIPNLRKAEQPKQPTAKPKAETAS 276
Db 76 ----TVEEIVAPSVVSVAPSVEEVE-----ENVEESVAENVEESVAENVEESVA 222
Qy 277 MPSEPSKQATVEKPKPEAKVAPE-----AKAEKPAVRPEPVPAAATASETAAESAP 331
Db 123 ENVEESVAENVEESVAENVEEIVAPTVEEIVAPSVVSVAPSVEEVEENVE 182
Qy 332 QEAASADTPTDGTGNVSEPVQVSABEETESGLFGSGSYTLILLAGGAALIALLLILR 391
Db 183 ESWAENVEESVAENVEESVAENVEESVAENVEES-----ENVEESVAENVEESVA 216
Qy 392 LAQSKRRRTTESVPPEE--EPDLDDAADGDIETTFAEVETPATPEPAPKNDVNDTLALDGE 450
Db 217 -----VAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEE 271
Qy 451 SEELSAKQTFVETDTPSNRIDLDFDLSAAQNGILSGALTQDEETQKRADADWNAIES 510
Db 272 SVEE-----NVEESVAEN-----VEESVAENVEESVAENVEESVAEN 308
Qy 511 TDSVVEPETFPNYP--VEIVDTPPEPSVAQTAENK--PETVDTDFSDNLPNNHIGTET 568
Db 309 VEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVEEVEENVEESVAENVEESVAENVEES 368
Qy 569 ASAKPASPLAGFLKASSPETILEKTVAEVQTPPELHDFLKVETDAVAETAPETPDNF 628
Db 369 V-----AENVEEIVAPTVEEIVAP--TVEEIVAPSVVSVAPSVEEVEEN 412
Qy 629 AAADLSALLQPAEAPSVSEENITETVAETPPFNATADLSALLQPSVPAVEENAAEIVA 688
Db 413 -VEESVAENVEESVAENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVA 466
Qy 689 DLSALLQPAEAPAVEENITETVAETSDFHTAADLSALLQPAEVPAPVEENVTKTVAEIP 748

Db 467 PSVSVSVAPSVEESVEENVEESVAENVE-----ESVAENVEESVAENVEESVAPTVEEI- 520
Qy 749 DFNATADDLSALLQPSVPAVEENAAE 775
Db 521 -----VAPSVEESVAPSVEESVAE 539

RESULT 11

US-10-205-331-66
; Sequence 66, Application US/10205331
; Publication No. US20040058326A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018199
; CURRENT APPLICATION NUMBER: US/10/205,331
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 66
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Microtubule associated protein IB
US-10-205-331-66

Query Match 6.1%; Score 269; DB 12; Length 2364;
Best Local Similarity 19.6%; Pred. No. 1.2e-07;
Matches 189; Conservative 131; Mismatches 283; Indels 362; Gaps 42;
Qy 94 TVSEKGLTAKVHKLGRKAVIA-----VSSEQAVRDPVLVFRIGAGAVREYTAILDPVGY 148
Db 505 SVTEKEVPSKEEQSPVKAEEAATEKSPKVTQKVVKEIKTKPEEKKEE----- 556
Qy 149 SPKTKSALSDOKTHRK--TAPTASQENQONAKALRKTDKDSANAAVK--PAYNGKTHTV 204
Db 557 KPKKEVAKKEDTKPLKDKKPKKEEAKKEIKKEIKKEEKELKKEKVKETPLKDAKKE--V 615
Qy 205 RKGETVKQIAAAIAPKHLTLFOVADALLKANPNVSAHGRRLRAGSVLHPIPNLRKAEQPK 264
Db 616 KDKK--KEVKEEKEPKKEIKKIDIKKSTP-----LSDIKK 652
Qy 265 PQTAKPKAETASMPSEPSKQATVE--KPVEKPEAKVAAPAEKPAKPAVRPEPVPAAATAA 322
Db 653 PAALKPK--VAKKEEPTKKEPIAAGKLKDKGKVKVKKKEGKKTTEAAA--TAGGTAATAA 707
Qy 323 SETAESAPO--EAAASADTPTDGTGNVSEPVQVSABE----- 361
Db 708 AAGVAASGPAKLEAERSIMSPEP-----LTKDFEELKAEEDIVAKDIKPOLKLEIEDEE 762
Qy 362 ---ETESGLFGSGSYTLILLAGGAALIALLLLRQAQSKRRRTTESVPEEPDLD---D 414
Db 763 KLKETEPG-----EAYVIQKETEVSQSAE 787
Qy 415 AADGDIETTFAEVETPATP--EPAPKNDVNDTLALD-----GSESEELSAKQ 459
Db 788 SPDEGITTTEGECEQTPPELEPVEKQGVDDIEKFDEGAGFERSAGDEEYEEKAETEE 847
Qy 460 TFDVETDTPSNRIDLDFDLSAAQNGILSGAL-----TQDEETQKRADADWNAIESTDSV 514
Db 848 AEEPEDEGEDN-----VSGSASKHSPTDEETIAK--AEADVHIKRRSV 890
Qy 515 -----YEPETFPNYPVEIVD-- 531
Db 891 ASGDDRAEDMDALEKGEAQSEBEGEEDKAEADAREEDHEPKTEAEDVMAVVDKA 950

[illegible]

Result No.	Query	Score	Match	Length	DB	ID	Description
							%
1	440A	100.0	880	2	G81786	Neisseria-specific	
2	4273.5	97.0	875	2	B81209	tspa protein NME03	
3	381.5	8.7	919	2	F83257	hypothetical prote	
4	353.5	8.0	1621	2	A82255	hypothetical prote	
5	316	7.2	1822	2	S33441	EF protein - Strept	
6	312	7.1	801	2	T29018	hypothetical prote	
7	276.5	6.3	1262	2	T29523	hypothetical prote	
8	272.5	6.2	1829	2	T24583	hypothetical prote	
9	271.5	6.2	1110	2	I51116	NF-180 - sea lamp	
10	269	6.1	2364	2	A56577	microtubule-associ	
11	266.5	6.1	1634	2	T26517	hypothetical prote	
12	264	6.0	2055	2	T31110	extracellular matr	
13	263.5	6.0	2464	1	D8WSPL	microtubule-associ	
14	260.5	5.9	880	2	D8S756	protein T23E7.2b [
15	258.5	5.9	5327	2	T13564	microtubule-associ	
16	257	5.8	1337	2	T30291	dextranase - Strept	
17	253.5	5.8	682	2	F83228	hypothetical prote	
18	253.5	5.8	3507	2	T34513	hypothetical prote	
19	253	5.7	1367	1	S48478	glucan 1,4-alpha-g	
20	248	5.6	873	2	A47283	calphotin - fruit	
21	248	5.6	971	2	T19431	hypothetical prote	
22	244.5	5.6	734	2	E42680	nucleolus-cytoplas	
23	244.5	5.6	2570	2	T17451	fimbriae-associated	
24	244	5.5	1229	2	T25697	hypothetical prote	
25	244	5.5	1616	2	G64242	Cyt adherence-acces	
26	243.5	5.5	1558	2	B71603	RESA-H3 antigen PP	
27	243	5.5	865	2	A47282	calcium-binding pr	
28	241	5.5	2187	2	T30826	nascent polypeptid	
29	240.5	5.5	1038	2	AG3187	hypothetical prote	

100

QY 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKQTFDVTDTSPNNRIDLDFDLSA 480
Db 421 EITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKQTFDVTDTSPNNRIDLDFDLSA 480
QY 481 AAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTTPESVAQ 540
Db 481 AAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTTPESVAQ 540
QY 541 TAENKPEVTDVDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKTVAEQ 600
Db 541 TAENKPEVTDVDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKTVAEQ 600
QY 601 TPEELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 660
Db 601 TPEELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 660
QY 661 NATADDLSALLQPAEAPSVENITETVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 720
Db 661 NATADDLSALLQPAEAPSVENITETVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETPDF 720
QY 721 ADDLSALLQPAEAPSVENITETVAETAPETPDFNATADDLSALLQPAEAPSVENITETVAETPDF 780
Db 721 ADDLSALLQPAEAPSVENITETVAETAPETPDFNATADDLSALLQPAEAPSVENITETVAETPDF 780
QY 781 PDSNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTFSPSESVDGSDAPSEAKYDLA 840
Db 781 PDSNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTFSPSESVDGSDAPSEAKYDLA 840
QY 841 EMYLEIGDRDAAAEVQKLEAEAGDVLKRAQALAEELGI 880
Db 841 EMYLEIGDRDAAAEVQKLEAEAGDVLKRAQALAEELGI 880

RESULT 2
C81209
tspA protein NMB0341 [imported] - Neisseria meningitidis (strain MC58 serogroup B)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81209
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Virology 287, 1809-1815, 2000
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81209
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-875 <TET>
A:Cross-references: GB:AE002391; GB:AE002098; NID:g7225561; PIDN:AAF40784.1; PID:g722556
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0341

Query Match 97.0%; Score 4273.5; DB 2; Length 875;
Best Local Similarity 97.1%; Pred. No. 1.6e-188;
Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

QY 1 MPAGLPRCPNMTKFTDCTRSNRIOPPHRGYIILKNNRQIKLIAASVAVASFOAHAGL 60
Db 1 MPAGLPRCPNMTKFTDCTRSNRIOPPHRGYIILKNNRQIKLIAASVAVASFOAHAGL 60
QY 61 GGLNLTQSNLDEFFSGSIITVGEAKALLGGGVTYSEKGLTAKVHLGDKAVIAYSSEQA 120
Db 61 GGLNLTQSNLDEFFSGSIITVGEAKALLGGGVTYSEKGLTAKVHLGDKAVIAYSSEQA 120
QY 121 VADPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDGKTHRKTAFTAESQENAKAL 180
Db 121 VADPVLVFRIGAGAQVREYTAILDPVGYSPKTKSALSDGKTHRKTAFTAESQENAKAL 180
QY 181 RXTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAARPKHLTLEQVADALLKANPNVSA 240

Db 181 RXTDKDSANAAVKPAYNGKTHTVRKGETVKQIAAARPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRLRAGSVLHIPLNLRIRKAEQPKPQATKPAKETAASMPSEPSKQATVEKPKVEKPAKVAA 300
Db 241 HGRLRAGSVLHIPLNLRIRKAEQPKPQATKPAKETAASMPSEPSKQATVEKPKVEKPAKVAA 300
QY 301 PEAKAEKPAVRPEVPAAANTAASATAESAPOEAAASAIIDTPTDETGNVSEPEVQVSAE 360
Db 301 PEAKAEKPAVRPEVPAAANTAASATAESAPOEAAASAIIDTPTDETGNVSEPEVQVSAE 360
QY 361 EETES-----GLFGGSYTLILAGGGAALIALLLLRQAQSKARTEESVPPEEDDLDDAA 416
Db 361 EETESGLFDGLFGGSYTLILAGGGAALIALLLLRQAQSKARTEESVPPEEDDLDDAA 420
QY 417 DDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKQTFDVTDTSPNNRIDLDF 476
Db 421 DDGIEITFAEVETPATPEPAPKNDVNDTLALDGESEBELSAKQTFDVTDTSPNNRIDLDF 480
QY 477 DSLAAAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTTPPE 536
Db 481 DSLAAAQNGILSGALTQDEETQKADADWNALESTDSVYEPETFPNPNPVEIVDTTPPE 540
QY 537 SVAOTAEKPKETVDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKT 596
Db 541 SVAOTAEKPKETVDTDFSDNLPNNHIGTEETASAKPASPGLAGFLKASSPETTILEKT 600
QY 597 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAE 656
Db 601 AEVQTPPELHDFLKVYETDAVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAE 660
QY 657 TPDFNATADDLSALLQPAEAPSVENITETVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETSD 716
Db 661 TPDFNATADDLSALLQPAEAPSVENITETVAETAPETPDFNAAADDLSALLQPAEAPSVENITETVAETPD 720
QY 717 FHTAADDLSALLQPAEAPSVENITETVAETAPETPDFNATADDLSALLQPAEAPSVENITETVAE 776
Db 721 FHTAADDLSALLQPAEAPSVENITETVAETAPETPDFNATADDLSALLQPAEAPSVENITETVAE 776
QY 777 TLETPDSDNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTFSPSESVDGSDAPSEAK 836
Db 772 TLETPDSDNTSEADALPDFLKDGEETVDMWSIYLSEENIPNNADTFSPSESVDGSDAPSEAK 831
QY 837 YDLAEMLYIEIGDRDAAAEVQKLEAEAGDVLKRAQALAEELGI 880
Db 832 YDLAEMLYIEIGDRDAAAEVQKLEAEAGDVLKRAQALAEELGI 875

RESULT 3
F83257
hypotheetical protein PA3115 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83257
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; L.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-919 <STO>
A:Cross-references: GB:AE004735; GB:AE004091; NID:g9949216; PIDN:AAG06503.1; GSPDB:GNOC
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3115

Query Match 8.7%; Score 381.5; DB 2; Length 919;
Best Local Similarity 23.0%; Pred. No. 3.6e-10;
Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;

QY 41 IKLIAASVAVASFOAHA-GLGGLNIQSNLDFPSSITVTGEEKALLGGGSVTVS--- 96

Db 8 VRAIAAA-SVLTSGMAGLGLGELTKSALNQPLDABIELL-EVRD-LGSGEVIPSLAS 63
 QY 97 -EKLGTAKVHKL-----GUKAVIAYSSQAVRDVPLVFRIGA---GAQVR 137
 Db 64 PEFSKAGVDRLLYLTDLKTPVVKNGKSVIRVTSSKPVQEPYLVFLVQVLPNGGLLR 123
 QY 138 EYTAILDVGVSPKTSKALSADGKTHRTAPTAESENQNAKALRKTDKDSANAAYK-- 194
 Db 124 EYTVLLDPPLYSFOAAASAQAPV---SAPRATC-----APRAPQAPVETTA 169
 QY 195 PAYNGKTHTVRKGETVKQIAAATRPKHLTLEQVADALLKANPNVSAHG---RLRAGSVLH 251
 Db 170 PAGSDTYRTV-SNDTLWEIAQRNTRDRVSPQAMLAFOELNPGAFVDFNGNIRLKSQGVLR 228
 QY 252 IPNLNRIKAPQPKPTAKPAETAAS-----MPSEPSK-QAT 286
 Db 229 IPTEQOMLERSPREALSQVQAQNSWREGSRNPAAGSAGARQLDATQRNAAGSAPSKVDAT 288
 QY 287 ----- 286
 Db 289 DNLRVLSGEGKAGKADKGGKGDKAIAADTLAVTKESLDSTRENEBELQSRMQLQSOLD 348
 QY 287 -VEKPEKPEAKVA-----APEAKAEKPAVRPEVPAAANTA 321
 Db 349 KUGKLQOLKDAQIAKLQGLGAGGOGAAQPNALPDASQPNAAQAAPAQPGTAAAAATP 408
 QY 322 ASETAESAPOEAAASAITDPTPE--TGNAVSPVPEQVSAEETESGLFGSYTLLAG- 378
 Db 409 APAGEAPAPAPQPPVAPPAPAEKPPAPAPAPAPVQAAEQAPAPFL-----DELLANP 463
 QY 379 -----GGAAIALILLLLLAQSKRARRTEESVP-----EBEPLDDAADGI-EITF 424
 Db 464 LMLAVIGGSALLALLVLLMLSRNAQKEEAQAFADTGEQEEDALDLGKGFDDTL 523
 QY 425 AEVETPATEPAKNDVNTLALDGESEEL-----SAQTEDVETDTPSNRIDLD 475
 Db 524 DPE-PQVAAVAPQVEKTTAQTSDALGEADIIYAYGRFQAELLQNAIYDEP-QRTDLR 581
 QY 476 FDSLAA-AQNGILSGALTQDEETQKRAADWNAIESTDSVYBETPNPNVPEIVIDTPE 534
 Db 582 LKLMVYAEWGREGFARGENELREIGGAQ-POVEQLKRY-----PAMVAAVA 631
 QY 535 PRESVAQTAENKPTVDTPDSNLPNNHIGTETASAKP-ASPSGIAGFLKASSPTEILE 593
 Db 632 GLAGAKLAQDELDSFSLD-DLSLDSGH-----AAKPDAAQDLDDAFDLSLDDGCG 683
 QY 594 KTVAEVQTE-ELHDFLVKVTDAVETAPETPDFAAAADDLSALLOPABAPSVEENITE 652
 Db 684 DVQADKSDSGALDDLTLDSLDLLAASTPADKP-----VDDLDFGLDFAE----- 728
 QY 653 TVAETPDPFNATADDLSALLOPSEVPAVERNAAEIVADDLSALLQ---PAEAPAVENVTE 709
 Db 729 -LAETPS-QPKHDDLGFSLDLADP--EDKLSG---DPLLSINDEVPAAPADNEFTLD 781
 QY 710 TVAETSDHEHTAAD-DLSALLQPAEVPAVEENVTKVAEIPDFNATADLSALLQFSEVP 767
 Db 782 TEAABEPALSLPDDFSLDADEFTPEAPAEKGEDSFAAQDLDEVSALQDELAS----- 833
 QY 768 AVEENAAEITLTTPDSNT---SEADALPDFLKGGEETVDWYIYSEENIPNNAOTSPPS 824
 Db 834 -----NLDEPKSATPSFAEDAASASALDGD-----ADDDDFD 866
 QY 825 ESVGSDAPSEAKYDLAEMYLETGDRAAEITVOKLLERAEGVVLKRAQA 873
 Db 867 LSGADEAAT-KLDLARAYIDMGDEGARDIIDEVL--AEGNDSQQA 911

RESULT 4

A82255

hypotheoretical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: A82255
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, J.; L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: A82255
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1621 <HEI>
 A:Cross-references: GB:AE004181; GB:AE003852; NID:99655454; PIDN:AAF94159.1; GSPDB:GN00.
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC0998
 A:Map position: 1

Query Match 8.0%; Score 353.5; DB 2; Length 1621;
 Best Local Similarity 23.1%; Pred. No. 1.5e-08;
 Matches 226; Conservative 129; Mismatches 316; Indels 307; Gaps 50;

QY 137 REYTAILDVGVSPKTSKALSADGKTHRTAPTAESENQNAKALRKTDKDSANAAYKPA 196
 Db 24 RFPQRLLPVAVMVTQTSTFVSAESIRLVGPDGQVQPTQ---YSENIVERNANN--EPG 78
 QY 197 -YNGKTHTVRKGETVKQIAAATRP-KHLTLEQVADALLKANP-----NVSAGHLRAGSVL 250
 Db 79 RFPGPTSA---NOTLMSIASQURPSSSVTVQTLAIYQLNPOAFENQNIH-TLIFGSL 134
 QY 251 HIPNLNRIKAEOPKPTAKPAETAASMPSEPSKQATVEKPVKEPE-AKVAAP--EKAEEK 307
 Db 135 RVESLAQISNSSTQDAVNIMASHQAKLNQTPD---TPVPVAPRPAPVATPKVEAVAT 191
 QY 308 PAVRPEVPAANTAASETAESAPOEAAASAITDPT----- 343
 Db 192 P---POVTP-----TAPOEKAPTELKTPAKPSQSTDAEVMALEEKNHTLRML 237
 QY 344 -----DETG--NAVSEPVQVSAEE-----ETESGLFGGSYTLLAGG----- 380
 Db 238 SQVQSEVSTLKGDELGNIRSEVERKLEERKAEASRLAPSALDNLISNGWLVALLA 237
 QY 381 -----AALIALILLRLIAQSKRARRTEESVPEEEDLDDAADGDIETFAEVETAPTE 434
 Db 298 LIPGLLIAIVLLLNRRSSAQOENPTQNNITSEMP-----TAAPVTLG 341
 QY 435 PAKKNDVNTLALDG-----ESEBELSAKQTF--DV-----ETDTPSN----- 470
 Db 342 PEQTEDIGDILLDDDLFTTDDKEENDAEKAFADEDDVFDADLNEDTDLNLDGQSDDL 401
 QY 471 -----RIDLDEPDSLAAQNGILSGALTQDE-----ETQKRA-----DADNNAI 508
 Db 402 FVGIDDDGDLDTFALNESANGI---SVNADDKALGLEEMERALNDVSEPTNDNLSF 458
 QY 509 ESTDSVYETFPENPNVPEIVIDTPEPE-----SVAQTAENKP----- 546
 Db 459 DIADENQMS-----DDIEALLSGDEENELLSDGKVDQSLDLDLLASELDALDDPAID 513
 QY 547 -ETVDTDFSDNLPNNHIGTETASAKPASPGLAG-----FLKA 585
 Db 514 TETLDTLLNDELASL-----SEEDDDDFDLGAGVAGDQDLDFASIEEQADLEQLEAKA 569
 QY 586 SSPETILEKTVAEVQTP-----FELHDFLVKVTDAVETA-----PETPFAAAA 631
 Db 570 IDETALLDEILAQAPLSEESTELLDELDFDKPENDEFDQAQTLQPEPTIDLEE 629
 QY 632 DDLALLQ-----PAEAPS---VEENITETV-----AETPDENATADDLSA 669
 Db 630 DSTQLLNEVLGPVPEELASGLIEDQNSTELLDELDDLDDESEIATEFSVAPEKLSV 689
 QY 670 -----LLOPSEVPAVERNAAEIVADD-----LSALLQPAEA-----PAVERN 707
 Db 690 EDGTFLFDELLEIEHQHPESASLPELATEDEFENSDTFIDLLNSAPAKDPLLEFVIDENE 749

QY 708 TETVARTSDPH-----AADDLSALLQPAEVP-----VEENVTKTV-----AE 746
 Db 750 AFAQADDFFNPTEGGLEDLS---QPSALPANEFCTPOEDWVFDDESSPTLEGNAE 806
 QY 747 IPFNATADLSALLQSEVPVAVENAEITLE---TPDNTSEA---DAL-PDFLKDGE 800
 Db 807 L-ELSSAEDDL-----PEQTATNETADELLAAQFQSNVTVDTSDDLAPDGLSQSVE 860
 QY 801 ETVDWS-IYLSEEN-IPNNAD---TSPPSGVSDAPSEAKYDLAEMYLEIGDRDAAE 854
 Db 861 EPLTNDLEIPEENDEPQALAEVTPSSAFDEQQVETEIEPSEPLAAASNDSEDLTALNE 920
 QY 855 TVQKLLBEAEGDVLKRAQ 872
 Db 921 L--DLPEYTEDVLADVQ 936

RESULT 5
 EF protein - Streptococcus suis
 C:Species: Streptococcus suis
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
 C:Accession: S33441
 R:Smith, H.B.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
 A:Description: Repeats in an extracellular protein of wek-pathogenic strains are absent
 A:Reference number: S33441
 A:Accession: S33441
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1822 <SMI>
 A:Cross-references: EMBL:X71880; NID:G298031; PIDN:CAA50714.1; PID:G298032

Query Match 7.2%; Score 316; DB 2; Length 1822;
 Best Local Similarity 19.9%; Pred. No. 8.9e-07;
 Matches 204; Conservative 165; Mismatches 363; Indels 294; Gaps 43;

QY 37 NNQIKLIAASV-----AVAAFPQAHAGLGLNIQSNLD--EPFSGSITVTGEEAKAL 87
 Db 682 DNERLKGIPDSFTVNSDGTVSVDYSA---GGVNVGDATDIKNAATNLADTRNQAKAE 737
 QY 88 LGGSVTVSEKGLTAK---VHKLGDKAVIASVSEQAVRDPVLVFRIGAGAQVRYTAIL 143
 Db 738 I-DTKLAHEKKAIEAKRDEAFSKIDDISLRAEQRAAKDAVAA---AAGDALKE---L 789
 QY 144 DPVGYSPK---TKSALSDGKTHRTAPTAESEONQAKALAKTKDKDSANAAV--- 193
 Db 790 DNKATEAKEKIDKAITASEINDAKNGEI--NLSABAVGEKAIQNAQKEKELAKAEVENK 847
 QY 194 -----KPAY-----NGKTHTVRK---GETVQIAAAIRPKHIT----- 223
 Db 848 AFEALEKYNPNLLEEEKKAYFDDIKESKEVAVEKINNAENTABITAAIDEAEIAYNED 907
 QY 224 -----LEQVAD---ALLKANPNVSAHRLRAGSVLHINLRIKAEQKQPTA 268
 Db 908 VINAQAQDALNLEKDESETKAIDANPNLTPEEKAKA-----IAKVEELVNNAESDILS 962
 QY 269 KPAKTASMPSPSQATVEKPEKPAKVAAPAEK---AEKPAVRPEVPVPAANTAASET 325
 Db 963 KPTPTVQAVEDKA-----DKOLAKVELOAAADGAKKGIENPNLTPEKDVAKKAVEDA 1017
 QY 326 AAESAPEQAAASAIID---TPTDETGNVSEPEVQVSAER--ETESGLFGGSVYTLLAGGG 380
 Db 1018 V-----KVATDAIDKASFT-EVDTATSDGVKAIADAEFFKATQKD----- 1056
 QY 381 AALIALLLLRLAQSKVARTTES-----VPEEPDLDADDGIEITFAEVETP 430
 Db 1057 -----AKNKIAKEAESAKKAIIDNPNLTPEKSAKNVAVEEAKVATAAIDKA 1104
 QY 431 ATPAPAPKNDVNDTLALDGSEBEELSAKQTFDVTETPT-----SNRIDLPDLSLAAQ 483
 Db 1105 STPD-----AVQVEEDKGVAAILNITAKADAKGVIAAKLADEIKKLEKQAEAE 1153

QY 484 NGILSGALTODEET-----QKRADAWNIAESTDSVVEPEFENPVPVEI----- 528
 Db 1154 KAIADASTMTNEEKAIKAKALQDVVDKGALEDAARVATNEIHEATTEKAKAEELAGEK 1213
 QY 529 -----VIDTPEPSVAQTAENKPTVDTFSDNLPFSNNHIGTETASAKPASPS 577
 Db 1214 SLTDTGKEARDVAVELAKELAKEAIRTEEBEATKIVEKLAEDTRKAIEDPNLSDSDKQ 1273
 QY 578 GLAGFLKASPETI-----LEKTVAEVQTPPELHDFLKVYETDVAETAPETPDFNAAA 631
 Db 1274 AEIKKLTDAVAKTULATIRDNADKRTQEAKEAQAALADLEKAKETQKIAD-----KAAI 1325
 QY 632 DDLSSALLQPAEAPSVEN-----ITEVTAETPDFN-----ATAD 665
 Db 1326 DRLTILYKGELEATKQDAKNKIADKAAAKAEATASNPNTLDAEKKTFTDAVDAEAVKAN 1385
 QY 666 D-LSALLQPEVPAVEENAAEIVADD-LSALLQPA-----EAPAVEE-----NVTE-- 709
 Db 1386 DAISAATSPADVQKEEDAGVAAIAEDVLDAAKQDAKNKIADKAAAKAEAGISNPNLTDAE 1445
 QY 710 ----TVAETSDFHTAADDLSALLQPAEVPVAVENVTKTVAIPDPFNATADDLSALLQPE 765
 Db 1446 KKTFTDAVDAEAVAKANDAISAATSPADVQKEEDAGVAAIAE-----DVLDAKQDAK 1497
 QY 766 VPAVEENAAEITLPTPSNTSEADALPDFLKDGEETVDWISILSEENIPNNADTSFPSE 825
 Db 1498 NKIAKES-----DAKSAIDANPN-LTDAEKESAKKAV-----DADAKAATD 1538
 QY 826 SV-GSDAPSEAKY-----DLAEMYLEIGDRDAAAEVQKL-----LEEAG 865
 Db 1539 AIDASTSPVEAQSAEDKGVGSIADVDLDAKQDAKNKIADKAAAKAEIDANPNLSDAEK 1598
 QY 866 DVLKRA 871
 Db 1599 EASKKA 1604

RESULT 6
 T29018
 hypothetical protein ZK84.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T29018
 R:Kirsten, J.
 A:Submitted to the EMBL Data Library, April 1995
 A:Description: The sequence of C. elegans cosmid ZK84.
 A:Reference number: Z20553
 A:Accession: T29018
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-801 <XIR>
 A:Cross-references: EMBL:U23181; PIDN:AAC48204.1; GSPDB:GN000020; CESP:ZK84.1
 A:Experimental source: strain Bristol N2; clone ZK84
 C:Genetics:
 A:Gene: CESP:ZK84.1
 A:Map position: 2
 A:Introns: 22/2; 45/3; 108/1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 7.1%; Score 312; DB 2; Length 801;
 Best Local Similarity 24.7%; Pred. No. 4.7e-07;
 Matches 205; Conservative 77; Mismatches 365; Indels 184; Gaps 33;
 QY 97 EKGLTAKVHKLGDKAVIASVSEQAVRDPVLVFRIGAGAOVR-----EYTAILDPVGYSPKT 152
 Db 77 ERDAAAAAKSTDAVNSASVDTGYCD-----GAVADVAAAPGPEPTPATAPEGAAEVP 129
 QY 153 KSAL-SDGKTHRTKPTAETAESEONQAKALRKTKDSDSANAQVKNVANGKTHTVRKGETVK 211
 Db 130 PTAVEGSGVSGGIP-----DEVATTVAADSGSDAPASONSVEITLTATEPAL 179
 QY 212 QTAAAIRPKHITLEQVADALLKANPNVSAHGRLAGSVLHINLRIKAEQKQPTAKPK 271

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1829 <WIL>

A:Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1

A:Experimental source: clone T06D8

C:Genetics:

A:Gene: CESP:T06D8.1

A:Map position: 2

A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match

Best Local Similarity 6.2%; Score 272.5; DB 2; Length 1829;

Matches 166; Conservative 121; Mismatches 344; Indels 183; Gaps 27;

QY 65 IQSNLDPEFGSITVTGEEKALLGGGVTVSEKGLTAKVHKLGDKAVIAVSSQOAVRDP 124

DB 726 VESDEBPASSSTSIPTLSK-----DQVTEAS-----GEETTAATASRETT 771

QY 125 VLVPRIAGAGQVREYTAIILDPVGYSPKTKSALSDDGKTHRKTAFTAESQENQNAKALRTD 184

DB 772 TSAVTEGSGEET-TVAVVSSGEEPPASSST-----SIPTLSKDDQVTEA---SG 818

QY 185 KDSANAAVKPAYNGKTHTVKGETVQIAAIPKPKHLTLBOVADALLKANPNVSAHRL 244

DB 819 EETTTAAATEASEETTTSAVTEGSGEDTTVVAV-----VSSGEG 858

QY 245 RAGSVLHPIPNLRKAEQPKQTAQKAEATASMPSEPSKQATVEKPV---KPEAKVAAP 301

DB 859 PASSSTSIPT---ELSKDDQVTEASGEETTTAAATEASEETTTSAVTEGSGEETTVAVV 915

QY 302 EAKAEKPAVRPEPAAANTAASATAESAPQEAASAIIDTPTDGTNAVSEPFVQVSAE 361

DB 916 ESSGEEPPASSSTSIPTLSKDDQVTEASGEETTTAAATEASEET---TTSVTEGSGEE 972

QY 362 ETESGLFGGSYTLILAGGGAALIALLLRLAQSKRARTTESVPEEPDLDDAADDGIE 421

DB 973 TTTSVTEGSG-----GEETTTSAVTEGSGEETTTAAATEASEETTTSAVTEGSGEET 1004

QY 422 ITFAVEPTATPEPAPKNDVNDTLALDGESEELSAAKOTFDVETDTPSNRDLDFSLAA 481

DB 1005 VTGSEIEIP-----SEESSSTTHD-----PSIPVITPKPSVSS 1039

QY 482 AQNGILSGALTQDEETOKRADWNAIBSTDSVYEPETFPNPNVPEIIVIDTPEPESVAQT 541

DB 1040 TIENVMS-KTSSERAAEKIIIGEHOQTKDDAGKEDD---NMPAFVTANPAGTSTTES 1094

QY 542 AENKPEVTDFDNLPSNNHIGTE-----ETASAKPASPGLAGFLK 584

DB 1095 AENVSTSGEED--ENIKMAKELGKQFAADLAKIAAKDGVNLTETADAKDSGETAHVEDEQ 1152

QY 585 ASSPETIL--EKTVAEV--QTPEELHDFLKVYETDAVA--ETAPETPDFNAADDLSALL 638

DB 1153 VSSSTESSGSEETTTVAKETTEEHE--ASGEDDAPAFVTCAPDTSITEASVTSALT 1211

QY 639 QPAEAPSVENITETVAETPDFNATADDLSALLQ---PSEVPVAEENAAEIVADDLSALL 695

DB 1212 DETTSSVADESTSGAGEVQSSAIIIDSATVASEEQTSSEATSVIESGSE-----1261

QY 696 QPAEAPVAEEN-VTETVAETSDPHTAADLSALLQPAEVPVAEENVTKTVAEIPDPNATA 754

DB 1262 ---EVTITDENIVTSTVAQ-----LEEGSGITAAESKEDSVT 1296

QY 755 DDLSALLQPSVEPVAEENAAETITLTPDSNTSEADALPDFLKDGEETVDMISYLEENI 814

DB 1297 TE--ATSGTTSVSSDGSGETVAPNDSETSTSSQSTTDDGSGVTAES--KDESS 1352

QY 815 PNNADTSPFSSVSGSDAPSEAKYDLAEMYLEIGD 848

DB 1353 TTEAPAFVTSKTSGGSEDEEDSPOTHEFLTIDE 1386

RESULT 9

151116

NF-180 - sea lamprey

C/Species: Petromyzon marinus (sea lamprey)

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C/Accession: I51116

R/Jacobs, A.J.; Kamholz, J.; Selzer, M.E.

Brain Res. Mol. Brain Res. 29, 43-52, 1995

A/Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation

A/Reference number: I51116; MUID:95287814; PMID:7770000

A/Accession: I51116

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1110 <NAC>

A:Cross-references: EMBL:U9361; NID:G632548; PIDN:AAA80106.1; PID:G632549

C:Superfamily: neurofilament triplet H protein

Query Match

Best Local Similarity 6.2%; Score 271.5; DB 2; Length 1110;

Matches 181; Conservative 119; Mismatches 330; Indels 219; Gaps 33;

QY 66 QSNLDPEFGSITVTGEEK-----ALLGGGVTVSEKGLT-----AKVHKLGDKAVIA 114

DB 177 QEHLDEEIQRLREKTDDEVLRLNETEALINAFRNKVVDDTSLVRMEMDKRTQSLLDEITFL 236

QY 115 VSSQAVRDPVLVPRIGAGQVREYTAIILDPVGYSPKTKSALSDDGKTHRKTAFTAES 171

DB 237 KKNHEEEVDLELL-----AQIQSTVSVERKDFAVPEITTAALREIRGLQEGSGARNIET 289

QY 172 QENQ-NAKALRKTKDSDANAAVKPAYNGKTHTVRK-----GETVQKQIAAA 216

DB 290 AEFWFKGFSQTEAAGQNDAIRSAKEEITHEHRKLMQRCCTELDALAGTKESLERQSLSE 349

QY 217 IRPKHLT-----LEQVADAL-----LKANPNVSAHGLRAG 247

DB 350 MEERHQSDVGNLQDAAQOENELNRTKWMARHLREYQDLLNVKMLDIEIAAARKLDDG 409

QY 248 SVLHPIPNLRKAEQPKQTAQKAEATASMPSEPSKQATVEKPV---EKPEAKVAAPAKA 305

DB 410 EEIHY-----SSGPLTPAKP-----PKAPSAKPAKAKVKVSKKAPKIKVES-----453

QY 306 EKPAVRPEPVA-ANTAASATAESAPQEAASAIIDTPTDGTNAVSEPFVQVSAEETE 364

DB 454 -----EPISAQDLDLEDLAQEEVWEAKAAPVVSAPKDEEEEEEEEEEEEEEEAE 506

QY 365 SGLFGGSYTLILAGGGAALIALLLRLAQSKRARTTESVPEEPDLDDAADDGIEITF 424

DB 507 E-----EEDRGKGEAE 535

QY 425 AEVETPATPEPAPKNDVNDTLALDGESEELSAAKOTFDVETDTPSNRDLDFSLAAQN 484

DB 536 AEVE-----EAEAEETAE 585

QY 485 GILSGALTQDEETOKRADWNAIBSTDSVYEPETFPNPNVPEIIVIDTPEPESVAQTAE 544

DB 586 --AAEAKAEVEEEAEAEAE--EEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 636

QY 545 KPETVDTDFSNLPSNNHIGTEETASAKPASPSGLAGFLKASSPTILEK---TVAEVQT 601

DB 637 EAEEAEAE-----EEVTSK-----KAKTQAEVVEEEAEAEAEAEAEAEAEAE 672

QY 602 PEELHDFLKVYETDAVETAPETPDFNAAADDLSALLQPAEAPSV-EENITETVAETPDF 660

DB 673 EEEAEAE--EAGEEDVEASEKEEEDDSKEAD---AEDEAEAEAEAEAEAEAEAEAEAE 727

QY 661 NATADDLSALLQPSVEPVAEENAAEIVADDLSALLQPAEAPVAEENVTKTVAEIS 720

DB 728 EA-----EEEAKESEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 773

QY 721 ADDLSALLQPAEVPVAEENVTKTVAE-----PDFNATADDLSALLQPSVEP 767

DB 774 SDD-----EKP--EEEVKSEAPVAEPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 823

QY 768 AVEENAAETITLTPDSNTSEADALPDFLKDGEETVDMISYLEENIPNNADTSPFSES 827

Db 824 KAWEVKKGAEPKPKAKPAKAAKKAEPVEKE-----EPEESFTBEPKPKAAK 875

QY 828 GSDAPSEAK 836

Db 876 PAKAPAKPK 884

RESULT 10

A56577

microtubule-associated protein MAP 1B - rat (fragment)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997

C;Accession: A56577

R;Rauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.

Eur. J. Cell Biol. 57, 66-74, 1992

A;Title: Identification of two distinct microtubule binding domains on recombinant rat

A;Reference number: A56577; MUID:92347374; PMID:1639092

A;Accession: A56577

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2364 <ZAU>

A;Cross-references: GB:X60550

A;Experimental source: brain

A;Note: nucleotide sequence not given; conceptual translation not complete

C;Superfamily: microtubule-associated protein MAP1B

Query Match 6.1%; Score 269; DB 2; Length 2364;

Best Local Similarity 19.6%; Pred. No. 0.00018;

Matches 189; Conservative 131; Mismatches 283; Indels 362; Gaps 42;

QY 94 TVSEKGLTAKVHKGDKAVTA-----VSSQAVRDPVLVFRIGAGAQVREYTAIDLPVGY 148

Db 505 SVTEKEVPSEKQSPVKAFAEAKAATESPKVTKDKVVKKEIKTKPEEKKE-----556

QY 149 SPKTKSALSDGKTHRK--TAPTAESQENQNAKALRTDKKDSANAIVK--PAYNGKTHV 204

Db 557 KPKVEVAKEDKTPLKDKKPKKBEAKETKEIKKEKKELKEVKKETPLDKAKKE-V 615

QY 205 RKGETVKQIAAARPKHLTLEOVADALLKANPNVSAHGRLAGSVLHINLNRKAEQPK 264

Db 616 KKDEK-KEVKEEKEPKKEIKKIDIKSTP-----LSDTKK 652

QY 265 PQTAKPKAETASMPSPSPQAVTE--KPVKEPEAKVAPEAKAEKPAVRPEPVAANTAA 322

Db 653 PAALPKP--VAKKEEPTKEPTAAGKDKGKVKVKEGKTEAAA--TAVGTAAVA 707

QY 323 SETAESAPO--EAAASALDPTDETGNVSEPEQVSAE-----361

Db 708 AAGVRAAGPAKLEAERSLMSFED-----LTNDPEELKAERIDVAKDIKQLELIEDRE 762

QY 362 ---ETESGLFGGSYTLILLAGGGAALIALLLLLLAQSKRAARTEESVPPEEPOLD---D 414

Db 763 KLKTEPG-----EAVVIQKEVEVSKGSAE 787

QY 415 ADDGIEITFAVEVTPATP--EPAPKNDVNDTLALD-----GESEBELSAKQ 459

Db 788 SPDEGITTTEGECEQOTPEEPEVKEKQGVDDIEKFDEGAGFERSSEAGDYVEKAETEE 847

QY 460 TFDVDTTPSNRIDLDFSLAAQNGILSGAL----TDEETQKRAADWNAIESTDSV 514

Db 848 AEPEDEDGDN-----VSGSASKHSPTDEETAK-AEADVHIKKEKRESV 890

QY 515 -----YEPETFNPNVPEIVID--531

Db 891 ASGDDRAEDMDALEKGAESQEEBEEEDKAEDAREDEHPKTEADYVMAVVK 950

QY 532 -----TPEESVAQTAEKNETVDVDFSNLPSNNH-----IGTEETASAKP 573

Db 951 AEAGVTEDQYDFLGTAPKQGVQSPSREPASSIHD--ETLPGSSSEATASDEENRQDP 1008

QY 574 ASPSGLAGFLKA-----SSPETILEKTVAEVQTPPELHDFLKV--YE-----613

Db 1009 EEFATSGTQSTIIEISSPTMDSEMSTPRDVMTDTNNEETSPESQSFVNITKYESLY 1068

QY 614 -----TDAVAETAPETPDFNAADDL-----SAL-----637

Db 1069 SOEYKPVVASFNGLSDGSKTDATDGRDYNASASTISPPSSMEEDKFSKALDAYRPEE 1128

QY 638 -----LQPAEAPSVSEENITETVAETP-----DFNATADDLSALLQOP 673

Db 1129 TDVKTGAELDIKDVSDERLSPAKSPSPSPPIEKTPLGERSVNFSLT-----P 1179

QY 674 SEVPVAEENNAETVAODLSALLQAPAPAVEENVETVAETSDFTHTAADLSALLQAPAEV 733

Db 1180 NEIKASAGEATAV-----VSPGVTQAVVEE--HCASPEEKTU-----EV 1217

QY 734 PAVEENVTKTVAIPDNATADDLSALLQSEVPVAEENNAETILETPODNTSEADALP- 792

Db 1218 VSPSQSVTSAGHTPYQSTDEKSSHL-PTFV-----TENAQAVPV 1258

QY 793 --DFLKGDEETVDMSIYLSEENIPNNADTSPFSESVGSD--AP-----SEAKYDLARMYL 844

Db 1259 SFEF-TEAKDENERSISPMDEPVP--DSESPLEKVLSPRLSPPLIGSEAY--EDFL 1311

QY 845 EIGDR 849

Db 1312 SADDK 1316

RESULT 11

T26517

hypothetical protein Y18D10A.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T26517

R;Harris, B.

submitted to the EMBL Data Library, December 1998

A;Reference number: Z20226

A;Accession: T26517

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1634 <WII>

A;Cross-references: EMBL:AL034393; FIDN:CAA22308.1; CESP:Y18D10A.1

A;Experimental source: clone Y18D10A

C;Genetics:

A;Gene: CESP:Y18D10A.1

A;Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

Query Match 6.1%; Score 266.5; DB 2; Length 1634;

Best Local Similarity 22.4%; Pred. No. 0.00014;

Matches 189; Conservative 105; Mismatches 310; Indels 239; Gaps 41;

QY 94 TVSEKGLTAKVHKGDKAVIAVSSQAVRDPVLVFRIGAGAQVREYTAIDLPVGY-YSPTK 152

Db 305 TVRRPDFTAKMTLELKAKS-----RAPRLV--ETKFKVKKW-----PEGIHKPTE 345

QY 153 KSA---LSDGKTHRTKAPTAESEONAKALRKTID--KKDSANAIVKPAYNGKTH-----202

Db 346 KDSFGLANSTKLRNQFPFISDPTETTAQTAERKALLGAGAGASEFGSSSSIIHKKGK 405

QY 203 TVR-----KGETVKQIAAAR-----PKHLTLEQVAD--ALLKANPN 237

Db 406 TVRELOTEMMKGESYKKAARMRAEASAGRSQAPGPAPAPAAASELQDPDFGLSMSDFG 465

QY 238 VSAHGLRAGSVLHINLNRKAEQPKQTAEPKAEATASMPSEPSQATVEKPVKEPEAK 297

Db 466 SDSETEEHQKQKHIPAM-----VTRSRALSLPVTPKASSSSKMPPPSPS 514

QY 298 VAAPEAKAEKP-----AVRPEVPVAANTAASETAESAPOEAAAASAIPTDETGNVSE 353

Db 515 PSTGRRGRRRTLTSMSEFAAAAVT-----PAPGRPRSRSAKVSENPESLSEAP 570

QY 354 VEQ-----VSAEETESGLFGGSYTLILLAGGGAALIALLLLLLAQSKRAARTEE 403

Db 571 VKRGGRPRSRSTWISDESPSTSS-----TAAKRSKRA-E 606

404	QY	SVPEEPDLDAADGCIETTFAEVETPATPEPAPKNDVNDTIALDGBSEELSAAQTDFV	463
607	Db	SDOEEEQDL-----KLTNKSPEKPKKPSXTTEETVGDVL-----KKRLRDTAKTTATV	654
464	QY	ETDTPSNRIDLDFDSLAAQNGILSGALTOBETQK-----ADADWNAIESTDS--VYE	516
655	Db	IHTPGPPLR-TRRWERMRAPT-----AVTSSKKKPKNAGSADGSINEEHEHETMILE	707
517	QY	PETFNPNPVEIVIDTPESPESVAQTAENKP-----ETVDTDFSNLPSNNHIGTEETA-	569
708	Db	EQT-----LDLPQ-----QTSQOQPRISCGSELLDEQD---ASEEHSCHVPSAP	749
570	QY	--SAKPASPSGLAGFLKASSPETILEKTVAEQTPEELHDFLKVETDAVAETAPETPDF	627
750	Db	ELTKNPAPPVPEASEASAPP-----KIDIPECATPIL-----ALALALPTVSP-	793
628	QY	NAAADDLSALLQPAEAPSPVEENTETVAETPDFNATADDLSALLQSPVPAVEENAAEIV	687
794	Db	-----TAL-----EPKQAKENTAELEPTTSEISGRAP--QALPTSSQPTTSGSAAPEV	840
688	QY	ADDLSALLQPAEAPAVEENTETVAETSDFHTAADDLSALLQPAEAPAVEENVTKTVAEI	747
841	Db	DLLEISLISGAK-----TTKT-----RKAAPVOKSISSTQQA	875
748	QY	PDFNATADDLSALLQSPVPAVEENAAEITLETPDNGTSEADALPDFLXGDEETVDWSI	807
876	Db	PPTSVQAPPTSC-----SAAFPVDDLSEILSGAKTTTKYTKTQMP-----PVDQKK	922
808	QY	YLSE-ENIPNADTSFPSESVGSDAPSEAKYDAEMY-LEIGRDRAAEIVQ-----KLLE	861
923	Db	ISSEAPPLSDSAPT-----SVHQOTPKSPQILNLSKYGLDISDSEDEEEBERGMEIVE	977
862	QY	EAE	864
978	Db	EEB	980

RESULT 12
T3110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C/Species: Abiotrophia defectiva
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T3110
R/Manganelli, R.; van de Rijn, I.
Infect. Immun. 67, 50-56, 1999
A/Title: Cloning and characterization of emb, a gene encoding the major adhesin
A/Reference number: Z20988; MUID:99081722; PMID:9864195
A/Accession: T3110
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2055 <MAN>
A/Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AA003320.1
C/Genetics:
A/Gene: emb

Query Match	6.0%	Score 264;	DB 2;	Length 2055;
Best Local Similarity	20.8%;	Pred. No. 0.00025;		
Matches 212;	Conservative 149;	Mismatches 393;	Indels 266;	Gaps 44;
Qy	15	KFTDCTRSNRIQPPTHRYGILKNNRQIKLIAASVAAAFQAHAGLGLNIQSNLD---	70	
Db	925	KTAGTTAININLP-----QGTQAKAIA--AIEAFAQKR-----LELQGRNDLUTE	969	
Qy	71	--EPPSGSITVTGEAK-----ALLGGSVTVSEKGLTAKVHKLGDKAVIAVSS-----EQ	119	
Db	970	ERNMALADLTAKAQAADQADVQNRNNTGVAGAKDNGV-AQLQGINPTAVVKPDPARNAIDQ	1028	
Qy	120	AVRPDPVLVFRITG-----AGA-----QVREYTAILDPPVGSVPKTKSALSDGKTHRKT	165	
Db	1029	AARDKEAEFFQANTKLTDEEKAATIKKVDAAARDAKAAIDRAGSGNDVNNVAVNQCKA----	1084	
Qy	166	APTAESQBNQAKAL--RKTDKCDSANAAVVKPAVNGK--THTVKGETVTKQIAAAAIKPKH	221	

1085	----	-A QAIKALDDSPSAKDTAKAAIQNAADAKKAAITANNALTOEKKAAATK----	1133
QY	222	L TLEQVADALLKANPVSA-----HGRLRAGSV--LH PNLNR K AEQPKPQT	267
Db	1135	-----QVEDEAAKAAQAAVDASRSKADVDRAKDOGLQKISDVPVOPPKLNAIAA VDQAA TD	1190
QY	268	AKPKAETASMPSEPSKQATVEKPEVK-----PEAKVAADPEAKAEKPAVRPEPV	315
Db	1191	KKAVINND TLTOEKEEAIRK VDEEAAKARQAINDATSNADVAAKQAGTQAINVPQT	1250
QY	316	PAANTAAS---ETAASAPQ-----EAAASAI D TPTDTGN AV-----	350
Db	1251	PAAKNAAKAAVEQAADAKKQAIENDPNLTROEKDAAIAKVDQETNKARQAI DAAT TNADV	1310
QY	351	-----S P VQVSAEETESGLFGGSV TLLAGGGAALIA L L L L RL LAQSKRARRTEESVP	406
Db	1311	TAKONEGTQAINAVPQTPK-----AKTDK NAV TQAAE DKK-----SAI	1349
QY	407	EEBP DD DAADOGIEITFA EVETPATPEPAPKNDVND TLALDGESEBELSAK TQF D VETD	466
Db	1350	ENDPNLTREEKDAK---AKVD EAT---KAKNA DAAT SNDD ETAKONEGTQAINAVPQ	1403
QY	467	TPSNRID LDPDS LAAQNG LISGA-----LTQD EETQK RAD D WNNA I E D SV	514
Db	1404	TPRAKTD-----AKNAV TQAADR KKDAI ENDPNLTREEKVAAKAKVDAEAKKAKDAI	1455
QY	515	YEBETFN-----P NP VEI D TPEPSV AQTAENK ETV D TDFSDNL	557
Db	1456	-DAATS NADV TAKONEGT KAIN DVPQTP---TAKTDK NAV TQAAAKKDAI EKD---	1506
QY	558	PSNNHIGTETASAKPASPGLAGFLKASSPETILEKTVAE VOTPBEL HD-----FL	609
Db	1507	---PNLTREEKDAK-----AKVDAEAKKAKDAI DAAT SNADVTA	1543
QY	610	KVYETD VA ETAPETPD FNAAAD LSALLQPAEA--PSVEE--N T--ETVAETPD PNAT	663
Db	1544	KONEGTKAIN DVPQTP--TAKTDK NAV TQAAADAKKDAI EKD PNLTREEKDAAKAKVDAE	1601
QY	664	A----D LSALLQ PSEVP AVEE---NAAEIV-----AD LSALLQPAEA--PAVE NV	707
Db	1602	AKKAKDAI DAATS NADV TQADK GKNAINAVPQTPETAKTDK NAV TQAAADAKKDAI END	1661
QY	708	TE TVAE-----TSD FHTAAD LSALLQPAE VPAVE NV T KVA S I P D--NATAD LS	758
Db	1662	NLTREEKDAKAKVDAEATKAKNAI DAAT SNADV TAKONEGT KAIN DVPQTP TAKTD AKN	1721
QY	759	ALLQ-PSEVP AVEENAAE ITLETPD SNTSEADALP DFLKDGE EETV DWS Y LSE N PNN	817
Db	1722	AVDQAA TDKKS AIENDPAL TREEKDAAKAVDAEATKAKNAI DAAT SNADV TQADK GKN	1781
QY	818	ADTSF PSES VGSDP SAEKYD LAEMYLEIG DRDAAE TVQK LLEAEGDVLKRAQALAE	877
Db	1782	A NAV PQTP T---AKTDK NAV DQ---AATDKKAAI ENDPAL TREEKDAAKAKVDAEAKK	1835

RESIT, T 13

RESULT 13
 QRMSP1
 microtubule-associated protein MAP1B - mouse
 N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated prote
 C:Species: Mus musculus (house mouse)
 C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 01-Sep-2000
 C:Accession: S07549; S44387; A33645
 R:Noble, M.; Lewis, S.A.; Cowan, N.J.
 J. Cell Biol. 109, 3367-3376, 1989
 A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contain
 A:Reference number: A33645; MID:90094539; PMID:2480963
 A:Accession: S07549
 A:Molecule type: mRNA
 A:Residues: 1-2464 <NO>
 A:Cross-references: ENMEL.X51396; NID:952989; PIDN:CAA35761.1; PID:G53000
 R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
 Arch. Biochem. Biophys. 310, 428-432, 1994
 A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.

A:Reference number: S44387; MUID:94234720; PMID:8179328
 A:Accession: S44387
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 653-663, 'IC', <SAN>
 C:Superfamily: microtubule-associated protein MAP1B
 C:Keywords: microtubule binding; phosphoprotein; tandem repeat
 F:589-786/Domain: microtubule binding #status experimental <MTB>
 F:589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-695
 R-K-E/D-X)
 F:1861-2064/Region: 17-residue repeats
 F:91,116,351,888,1124,1153,1158,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: pH
 F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (cd
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 6.0%; Score 263.5; DB 1; Length 2464;
 Best Local Similarity 19.3%; Pred. No. 0.00034;
 Matches 180; Conservative 132; Mismatches 302; Indels 319; Gaps 38;

QY 94 TVSEKGLTAKVHKLGDKAVIA-----VSEQAVRDPVLVFRIGAGAVREYTAILDPVGY 148
 DB 601 SVTEKEVSKSEQSPVKAIEAQAATESKPKVTKDKVVKKEI-----KTKLEEKKEE 652
 QY 149 SPKTKSALSDGKTHRK--TAPTAESQENQAKALRTDKDSANAIVK--PAYNGKHTV 204
 DB 653 KPAKEVVKKEDKTPKKDKPKRVERVKKEIKKIKKEERKELKKVKKGTPLKDAKKEV- 711
 QY 205 RKGETVQKIAAIRPKHLTLEQVADALLKANPNVSAHGRLAGSVLHIPLNRIKAEQPK 264
 DB 712 -KKEEKVKKEEKEPKKEIKKISKDKSTP-----QSDTKK 748
 QY 265 PQAKPKAETASMPSPSKQATVEKV-----EKPEAKVAPEAKAEKPAVRPEVPAA 318
 DB 749 PSALKPKV-----AKKEESTKKEPLAAGKLKDKGVKIKKEGKTTTAA-----ATAV 796
 QY 319 NTAASETA-----AESAP--QERAAASAIPTPTDETGNVSEPVQVSABEETSGLF 368
 DB 797 GTAATTAUVAAAGIAASGPVKELEAERSLMSPEDLTDKFELKAAEEDVAKDIKP-- 853
 QY 369 GGSYTLILAGGAALIALLLRLAQSCKRARTSESVPEEPDLD-----DAADDGILEIF 424
 DB 854 -----QLELIEDEKELKQTEGAEVIOKETEVSKGSAESDEGITTTT 897
 QY 425 AEVETPATP--EPAPKNDVNTLALD-----GESEELSAKQTFDVEDTDS 469
 DB 898 GECECEQTPPELEPVKQGVDDIEKPEDRGAGFERSETGDVEEKAETEEAEEPEDEG 957
 QY 470 NRILDFSLAAQNGILSGALTQDEETOKRADADWNAIESTDSV----- 514
 DB 958 N-----ASGSASHSPTEDDSAK--AEADVHLKEREVSVDGDRAEEDMDVYL 1005
 QY 515 -----YEPETFPNPNVEIVD-----TPPESVAQTA 542
 DB 1006 EKGAEEQSEEGEEDKAEDAREEGYDPKTAEDYVMAVADAKAAEAGVTEQYGLGTS 1065
 QY 543 ENKP-----ETVDTDFSDNLPNNH-----IGTEETASAKPASPLAGFLKA----- 585
 DB 1066 AKQPGIQSPREFASSIHDETLPGGSESEATASDENREDQEEPTAUSGYTQSTIEISS 1125
 QY 586 -----SSPETILEKVAEQTEELHDFLKV-----YETDAVAE----- 619
 DB 1126 EPTFMDWSTPRDVMSTDNNEETESPSQEFNITKYESSLYSQEYSKPAVASFNGLSEG 1185
 QY 620 ---TAPETPDFNAADDL-----SAL----- 637
 DB 1186 SKTATDCKQYNASASTISPPSSMBEDKFSKALRDAYCSEKELKASAEIDIKOVSDER 1245
 QY 638 LQPAEAPSVEENITETVETP-----DFNATADLLSALLQSPSEVPVAVEENAAEIVADDL 691
 DB 1246 LSPAKSPSLSPSPPIEKTPLGERSVNFSLT-----PNEIKVSNAE----- 1286
 QY 692 SALLQPAEAPAVEENVTETVARTSFHTAADDLSALLQPAEVPVAVEENVTETVABIPDN 751

Db 1287 -----GEARSVSPGVQAVVEE---HCASPEEKL-----EVVSPSOSVGTSGAGTTPYQ 1333
 QY 752 ATADDLSALLQSPSEVPVAVEENAAEITLETDSNTSEADALPDFLKGEBETVDWSIVLSE 811
 Db 1334 SPTEKSSHL-PTE---VSENAQAVP---SFFPSEA-----KQENERA---SLSPMD 1376
 QY 812 ENIPNNADTSFPSESV-----GSDAPSE 834
 Db 1377 EPVP---DSESPVEKVLSPURSPPLIGSRSPYE 1406

RESULT 14
 D89756
 protein T23E7.2b [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 R:Accession: D89756
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: D89756
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-880 <STO>
 A:Cross-references: GB:chr_X; PIDN:ABM71258.1; PID:G24335547; GSPDB:GN000028; CESP:T23E7.2
 C:Genetics:
 A:Gene: T23E7.2b
 A:Map position: X

Query Match 5.9%; Score 260.5; DB 2; Length 880;
 Best Local Similarity 22.0%; Pred. No. 0.00012;
 Matches 174; Conservative 89; Mismatches 284; Indels 243; Gaps 36;

QY 172 QENQAKALRTKD-----KDSANAIVKPAYNGKTHTVRKGETVKQI-----AAAIRPK 220
 Db 31 QDSNAQLVEGGDQLVTAEDDEKSA-----QQVEDRTEDQVDPDITGTETVDDPM 81
 QY 221 HUTLEQVADALLKANPNVSAHGRLAGSVLHIPLNRIKAEQPKQTAKPKAETASMPSE 280
 Db 82 TTSMGIED--LPANDEAQVE-----ESAPADEPAPEEEKQPEAEPEEA 125
 QY 281 PSKQATVEK-PYEKPEAKVAAP--PAKAEKPAVRPEVPVPAANTAASETAAESAPQAAAA 337
 Db 126 PAEEAPQBEVPABEVPAAEETAPEAAVEEPPVE---EAPVEEPAEAEETAPAE--EP 179
 QY 338 AIDTPTDETGNVSEPVQVSAEETEGSLFGSGVTLILAGGGAALIALLLRLAQSQR 397
 Db 180 LIEVAVEA--PAEFPVAE--EAEPEAKQEF-----LQAPPEGR 217
 QY 398 ARR-----TESVPEEPDLDLDAADGIRI-----TPAEVETPATPEP 435
 Db 218 PRTPNDLNPNTADEPAAAEADAGSEVPPTPKTAASEKQSVPTPKSAAPSEVPPTPKS 277
 QY 436 APKNDVNTLALDGESEELSAKQTFDVEY-----DTPSNRIDLF-----DS 478
 Db 278 AAPSEVPPTKSAAPSEVPLTPKSAQSVPTPKSAVPSEAPPTPERSAAPSVDVPTPKS 337
 QY 479 LAAQNGILSGALTODEETOKRADADWNAIEST-----DSVVE----- 516
 Db 338 NGHANGSIPTGPSVAQSAKATPAATPAATPAATPAATPAATPAATPAATPAATPAATPAAT 397
 QY 517 PETFPNPNPVEIVDTPEPESVAQTAENKPTETVDTDFSDNLPNNHICTEETASAKPASP 576
 Db 398 PRSSVPATPTESNLTTPAKT-----PKTPKTPPTPTPKT-----PPTP 437
 QY 577 SGLAGFLKASSPETILEKTVAEVQ--TPBELHDFLKVYETDVAETAPETPDNFNAADLLS 635
 Db 438 -----KTPAVVEPEPEPVAE-----EPEPVAEPEPE--PEPVAE----- 470
 QY 636 ALLQPAEAPSVEENITETVARTPDNFNATADDLSALLQSPSEVPVAVEENAAEIVADDLSALL 695

```

471  ---EPAEPAVEE---PAEPE-----PADEATEPTAE----- 501
696 QPAAEPAVENVTET-VAETSDPHTAADDLSALLOPA-----EYPAVENTKTVAEIPDF 750
502  --AEPEAVEESIEKTEVVEESAPPAARQSPSPARRRQSPSPERQRTSRHAD-RDI 558
751 NATADDLSALLOQSEVPAVENAETIETPDNNTSEADALPOFLKDGEB---ETVDWSI 807
559 TSYDEDSYRAVPPRMP-----TATFSWSPPDKQSYTPISP-FVSTANKYRNEYTSGSS 613
808 YLSEENIPNADTSFSPESVGDSPSEAKYDLAEYMLEIGDRDAAAEVQKLAEEARGDV 867
614 YR-----PTNMYTSHFDDIVATGAFSSALY-----STNRLIERSRST 651
868 LKRAQALAE 877
652 RERQAMRSQ 661

RESULT 15
Ti3564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 5.9%; Score 258.5; DB 2; Length 5327;
Best Local Similarity 18.8%; Pred. No. 0.0016;
Matches 173; Conservative 144; Mismatches 325; Indels 277; Gaps 33;

QY 66 QSNLDPEPFGSTVTGGEAKALLGGGVTVSEKGLTAKVHKLGDKAVIAVSQAVRDPV 125
DB 1387 ETSRPESATGSKVTEQTKS-----KSPVPSPESEAKDKK 1424
QY 126 LVFRIGAGQVREYTAILDPCVSPKTKSALSDGKTHRTAPTAESENQNAKALRTDK 185
DB 1425 SPFASGEASR-----PESVAESVDEAGKASRRSIAKTHKDESSLDKAKEQSR 1475
QY 186 KDSANAAVKPAYNGKTHTVRKGETVKQIAA---AIRPKHLT-----LEQVADALLKA 234
DB 1476 RESLAESIKP-----ESGIDKESALASKEASRPESVTDKSKEPSRSRESIAESL--- 1523
QY 235 NPNVSAHGRLAGSVIHIPNLNRIIAEQPKQPTAKPKAETASMP-----SE 280
DB 1524 -----KAESTKDEKSPAPPSPGVSRRPGSVVSVESVKDETEKSKE 1559
QY 281 PSKQATV---EKP-----VEKPEAKV-----AAPEAKAEKP-----AV 310
DB 1560 PSRRSIAESAKPTEFREVSRPESVIDGIDKESAKPESRRDSPSLASKEASRPESVLESV 1619
QY 311 RPEPVAANTAASATAE-----SAPQEAASAIPTPTDETG---N 348
DB 1620 KDEPIKSTKSRRESVAESFKADSTKDEKSPKTSKDISRPSGAVENVMAPPFKTSRPES 1679
QY 349 AV-SEPVOVSAEETESLFGGSYTLILLAGGAALIALLLLRLAOSKXAERTTESVPE 407
DB 1680 AVGSMKDESMSEKPSRSVKDGA-----AOSRETSR-PASVAE 1717

```

```

QY 408 EEPDLDAADDGIIETFAEVETP-----ATPEPAKNDVNTLADGSESE 453
DB 1718 ---SANDGADDLKELSRPESTTQKEAGSIKDEKSPLASSEASRPASVAESVKDEAEKSK 1774
QY 454 ELKAKQTFVETTPSNRIOLDPDSLAAQNGILSGALTODEETOKRADADWNAIEGSTD 513
DB 1775 EESRRRESVAESKSPLSK---EASRPASVAESIKDEAEKSKESRRRESVAESKSPLS--- 1827
QY 514 VYEPETPNPNVEIVDTPEP-----ESVAQTA-----ENKPTVDTDFSDNLP 560
DB 1828 ---KEASRPASVAESIKDEAEKSKESRRRESVAESKSPLSKEASRPASVAESIKDEAEK 1884
QY 561 NHIGTEFTASAKPASPSGLAGFLKASSPETILEKTVAEVQTPPELHDFLKYETDAVET 620
DB 1885 KEESRRRESVAESKSPLSK---KEASRPASVAESIKDEAEKSK-----ESRESVAEK 1933
QY 621 AP-ETPDFNAAADDLSALLOPAEAPSVENITETVAETPDFNATADDLSALLOQSEVPAV 679
DB 1934 SPLPSKEASRPASVAESIKDEAE-KSKEESRRRESVAE-----KSPLPSK 1976
QY 680 EENAAEIVADDLSALLOPAEAPAVEENVETVAETSDPHTAADDLSALLOQPAFVPAVEEN 739
DB 1977 EASRPASVAESIKDEAEKSK-----EESRRRESVAESK-----PLPSKEAS 2016
QY 740 VTKTVAEIPDFNATADDLSALLOQSEVPAVEENAAEITLETDPDNTSEADALPDLKGE 799
DB 2017 RPASVAE-----SIKDEAEKSKESRRRESVAESK-----PLPSKEASRPASVAESIKDEA 2066
QY 800 EETVDMGIYILSEENIPNADTSFSPESVGDSPSEAKYDLAEYMLEIGDRDAAAEVQKL 859
DB 2067 EKS-----KEESRRRESVAESKSPLSKEASRP-----SVAESVKDEADKSKESRRRESM 2115
QY 860 LEEAEGDVLKRAQALAE 878
DB 2116 AESGKAQSIKGDQSPLEK 2134

```

Search completed: September 2, 2004, 19:21:37
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:11:46 ; Search time 27 Seconds
(without alignments)
1697.101 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MFAGRLPRCPMMTKFTDCT.....EEAGDVLKRAQALQELGI 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	276.5	6.3	2468	1	MAPB_HUMAN
2	269	6.1	2459	1	MAPB_RAT
3	263.5	6.0	2464	1	MAPB_MOUSE
4	257	5.8	1337	1	DEXT_STRDO
5	253	5.7	1367	1	AMVH_YEAST
6	244	5.5	1616	1	P200_MYCGE
7	243	5.5	865	1	CPN_DROME
8	240	5.4	704	1	NP14_RAT
9	239	5.4	1664	1	SLPI_CLOTM
10	238	5.4	3664	1	MINT_HUMAN
11	236	5.4	6632	1	UNB9_CABEL
12	232.5	5.3	1849	1	IGA4_HAEIN
13	221.5	5.0	1018	1	HMW1_MYCPN
14	221	5.0	1087	1	NFH_MOUSE
15	220.5	5.0	831	1	NFH_RAT
16	219.5	5.0	8545	1	ANCL_CABEL
17	219	5.0	1781	1	AKL2_HUMAN
18	219	5.0	2004	1	MY83_HUMAN
19	218	5.0	1026	1	NFH_HUMAN
20	215.5	4.9	1159	1	N124_SCHPO
21	215	4.9	1694	1	IGA2_HAEIN
22	214	4.9	1233	1	YF16_YEAST
23	214	4.9	3644	1	MINT_MOUSE
24	213	4.8	1702	1	IGA2_HAEIN
25	210.5	4.8	442	1	BNB_DROME
26	210.5	4.8	1238	1	SBCC_RHOCA
27	210	4.8	3584	1	NBEA_DROME
28	209.5	4.8	3924	1	ANK2_HUMAN
29	209	4.7	1192	1	RTN4_HUMAN
30	207.5	4.7	1411	1	TCOF_HUMAN
31	206	4.7	5147	1	PCLO_HUMAN
32	205	4.7	1861	1	MAP2_RAT
33	202.5	4.6	1385	1	FAT1_SCHPO

34	202	4.6	705	1	ICAL_BOVIN
35	202	4.6	5085	1	PCLO_RAT
36	201.5	4.6	2090	1	N214_HUMAN
37	201.5	4.6	2805	1	MAPA_HUMAN
38	201.5	4.6	3381	1	PGCV_BOVIN
39	201	4.6	778	1	HTR6_HALN1
40	200.5	4.6	1690	1	C190_DROME
41	200	4.5	670	1	VG50_HSV11
42	200	4.5	2774	1	MAPA_RAT
43	197	4.5	1262	1	STNE_DROME
44	197	4.5	1972	1	P531_HUMAN
45	197	4.5	5120	1	PCLO_CHICK

ALIGNMENTS

RESULT 1
MAPB_HUMAN
ID MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE LCL]
DE MAP1B.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.";
RL Genomics 22:273-280(1994).
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LCL1, LCL2 and LCL3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LCL1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -!- SIMILARITY: TO MAP1A.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L06237; AAA18904.1; -
Genew; HGNC:6836; MAP1B.
MIM; 157129; -
GO; GO:0005875; C:microtubule associated complex; TAS.
InterPro; IPR000102; MAP1B neuraxin.
Pfam; PF00414; MAP1B neuraxin; 10.
PROSITE; PS00230; MAP1B NEURAXIN; 6.
Microtubule; Repeat; Phosphorylation.
CHAIN ? 2468 MAP1 LIGHT CHAIN LCL1.
FT

FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;
Query Match 6.3%; Score 276.5; DB 1; Length 2468;
Best Local Similarity 20.1%; Pred. No. 0.00051;
Matches 185; Conservative 125; Mismatches 333; Indels 277; Gaps 35;
QY 150 PKTSALSDGKT--HRKTAPTAEQENQAKALRTDKDSANAAVKPAYNGKTHTVRK 207
DB 656 PKKEVAKKEDTKPIKKEKPKKEEVKKEVKEKKEKPKKEVKEKKEKPKKEVKEK 715
QY 208 ETVKQIAAIR-PKHLTLEQVADALLKANPNVSAHGRIRAGSVLHIPNLNRKAKQPK 266
DB 716 EKEVKEKKEPKKEIKKLPDANKSSTP-----LSEAKKPA 753
QY 267 TAKPKAETASMPSESKQATVE-----KPKVEKPAKVAAPKAEKPAVRPEVPAA 319
DB 754 ALKPKV-----PKKESVKKDSVAAGKPKKEKIKVKEKGAAB-AVAAVGTGAT 804
QY 320 TAASETAAESA-----POEAASAIDTPTDTGNVSEPVQVSAABETESGLFGGS 373
DB 805 TAAVMAAGAGIAIGPAKLEAERSLSMSPEDLTQPEELKABEVDVTDIKP----- 856
QY 374 LLLAGGAALIALLLLRQAQKRARRTEESVPEEPDLD-----DAADGDIETAEVET 429
DB 857 -----QLELIEDEKLKTEPVVYVQKEREVTKGPASPDGIGITTEGECEC 905
QY 430 PATP---BPAPKNDVNDTAL-----DGE----- 450
DB 906 EOTPELEPVEKQGVDDIEKFDEGAGFESSETGDYEBEKAETEAEBEEDGEEHVCVS 965
QY 451 -----SEBELSAKQTFDV-----ETDTPSNRIDLDFSLAAQNGILSGALQDET 497
DB 966 ASKHSPTDEESAKAEADAVIREKRSVAGSDRAEDMDE-AIEKG-----EAEOS 1016
QY 498 QKRADADNAISTDSVPEPFNPNYVEIIVD-----TPPEPSVAOT 541
DB 1017 EBEADEEDKAEDAREEYEPKMEAEYVMAVVDKAAEAGGAEEOQVGLTTPTKQLGAOS 1076
QY 542 AENKPTVDTDFSDNLPNNH-----IGTEETASAKPASPGLAGFLKA----- 585
DB 1077 PGREPASSIHD---ETLPGGSESEATASDENREDQEEFTATSGYTSQSTIEISSETPMD 1134
QY 586 --SSPETILEKTVAEVQTPPELHDFLKV--YETDAVAE---TAPETP----- 625
DB 1135 EMSTPRDVMSETNNEETSPSQEFVNIUKYSSLSYQRYSKPADVTPLNGFSEGSKTDA 1194
QY 626 -----DFNAAA-----DQLSAL-----LOPA 641
DB 1195 TDGKOYNASASTISPPSSMBEEDKFSRSLRDVACSEVKASTTLDIKDSISAVSEKVS 1254
QY 642 EAPSVENITETVAETP-----DFNATADDLSALLQPSVEPVAEENAEIVADD-LSAL 694
DB 1255 KFSLSFSPSPLEKTPLGERSVNFSLTPNEIKVSAEAEVAPVSPVETQVVEEHCA 1314
QY 695 LQPAEAPAVENVTETVAETSDPHTAADDLSALLQPAV-----PAV-----EENV 740
DB 1315 DKLEVVPSQSVTSAGHTPYQSPDTEKSSHL-PTEVIEKPPAVPVPVSPFSDAKDENE 1373
QY 741 TKTVA-----EIPDNATADDLSALLQPSVEPVAE--BENAAEITLETSDNTSEADALPDF 794
DB 1374 RASVSPMDPEVPD-----SESPIEKVILSPRSPPLIGSESAYESFLSADDKASGRCAESFFE 1430

QY 795 LKQGEETVWSYLSLENIPNADTSPSPESVSGSDAPSEAKYDLAEWYLEIGDR-----D 850
DB 1431 EKSGKQSPDQVSEVSE-----MTSTLSYQDKQEKSTDFAPIKEDFGQEKKTDD 1480
QY 851 AAATVQKLL---EEAEGDV 867
DB 1481 VEAWSSQPALALDERKLDGV 1500
RESULT 2
MAPB RAT STANDARD; PRT: 2459 AA.
ID MAPB RAT P15205; Q62958; Q9ER21; Q9QW92;
AC 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN-Sprague-Dawley; TISSUE-Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
RT recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rientz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
RT nervous system that is immunologically related to microtubule-
RT associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -!- INDUCTION: By nerve growth factor.

or send an email to license@isb-sib.ch.

```
CC -----
DR EMBL; Z38061; CAA8176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CAA32069.1; -
DR PIR; S48478; S48478.
DR GarmOnline; 139731; -
DR SGD; S0001459; MUC1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0030447; P:filamentous growth; IDA.
DR GO; GO:0007125; P:invasive growth; IMP.
DR GO; GO:0007124; P:pseudohyphal growth; IMP.
DR GO; GO:0007124; P:pseudohyphal growth; IMP.
DR Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
FT SEQUENCE 1367 AA; 91C00E2DBD61AA9D CRC64;
SQ
Query Match 5.7%; Score 253; DB 1; Length 1367;
Best Local Similarity 18.2%; Pred. No. 0.0023;
Matches 132; Conservative 130; Mismatches 398; Indels 64; Gaps 17;
QY 152 TKALSDDGTHKRTAPTAEQSE-----NONAKALRKTOKDSANAAPVYNGKTHT 203
DB 208 TKSTSTSTSESSSTSTSTSESSSTSTSTSESSSTSTSTSESSSTSTSTSTSTSTST 267
QY 204 VRKGTVKQIAAARPKHLTLQEVADALLKAMPNVSAHGRLAGSVLHPIPNLNRKABQ- 262
DB 268 TKERTPTPTTSCKEKTPPHDDTPTCKKTTTSCTKTTPVPTPSSSTSESSA 327
QY 263 PKPQAKPAEATAMSPSPSKQATVEKVPKPAEAPAEKAEKPAEPAEPAEPAEPAE 322
DB 328 PVPTPSSSTSESSAPVTSSTSESSAPVPTPSSSTSESS-----APVTSSTSES 378
QY 323 SETAESAPQEAASAIPTDPTGNAVSEPVQVSEAEETESGLFGGSYTLALLAGGAA 382
DB 379 SSAPVTSSTSESSAPVPTPSSSTSESSAPV--TSTSESSAPVTSST--ESSAP 433
QY 383 LIALLLLRLOAKRARETEEVPEEPDLDAADDGIEITPAEVEETPAEPAEPAEPAE 442
DB 434 VTSSSTSESSAPVTSSTSESSAPVPTPSSSTSESSAPVTSSTSESSAPVPTPSSST 493
QY 443 DFLA--LQGESEELSAKQTFVETDTPNSRDLDFDSLAQAQNGLSGALQDEETQKR 500
DB 494 ESSAPVTSSTSESSAPVPTPSSSTSESSAPVPTPSSSTSESS--SAPVTSSTSESS 551
QY 501 ADADWNAIESTDSVVEPETFNPNVEIIVDTPEPEVAQTAENKPEVTDVDFSNLPSN 560
DB 552 APVPTPSSSTSESSSTPTVTSSTSESSAPVPTPS-----SSTSESSAPVPTPSSST 607
QY 561 N-----HGTETASAKPASPGLAGFLKASP-ETILEKTVAEVQTPPEELHDFLKVE 613
DB 608 SAPAPVTSSTSESSAPVTSSTSESSAPVPTPSSSTSESSAPVPTPSS--STSES 663
QY 614 TVDAVETAPETPDFAAADDLSALLQPAEAPSEVEENITVETATPDFAATADDLSALL 673
DB 664 SSAPVPTPSSSTSESSAPVTSSTSESSAPVTSSTSESSAPVPTPSSSTSESSAPV 723
QY 674 SEVPVAEENAEIVADDLSALLQPAEAPSEVEENITVETATPDFAATADDLSALLQ 730
DB 724 TPSSSTSESSAPVPTPSSSTSESSAPVTSSTSESSAPVPTPSSSTSESSAPVPTPS 783
QY 731 -----AEVPVAEENVTK--TVAEIPDFNATADDLSALLQPEVP-AVEENAEITLET 780
DB 784 SSTSESSAPVPTPSSSTSESSAPVPTPSSSNITSS--APSTPSSSTSESSVPVPT 841
QY 781 PDSNTSEADLPFLKDGRETVDSIYLSEENIPNADTSFP-----SESVGSD-APS 833
DB 842 PSSSTSESSAPVSSSTSESSAPVPTPSSSNITSSAPSSIPFSSSTSESSFGTIVTTPS 901
```

```
QY 834 EAKY 837
DB 902 SSKY 905
RESULT 6
P200 MYCGE STANDARD; PRT; 1616 AA.
ID AC Q49429; Q49259; Q49298; Q49352; Q49353;
AC Q49429; Q49259; Q49298; Q49352; Q49353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein P200.
DE MG386.
GN Mycoplasma genitalium.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.F., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal genome complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN
CYTADHERENCE (BY SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39720; AAC71613.1; -
DR EMBL; U02245; AAA03400.1; -
DR EMBL; U02245; -; NOT ANNOTATED CDS.
DR EMBL; U02175; AAD12458.1; -
DR EMBL; U02136; AAD12402.1; -
DR PIR; G64242; G64242.
DR TIGR; MG386; -
KW Cytoadherence; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389
FT REPEAT 1205 1236 1-1.
FT REPEAT 1358 1389 1-1.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT REPEAT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;
Query Match 5.5%; Score 244; DB 1; Length 1616;
Best Local Similarity 20.0%; Pred. No. 0.0065;
Matches 178; Conservative 124; Mismatches 301; Indels 286; Gaps 43;
QY 92 SVTVSSKGLTKVY-----HKLGRKAV-IAVSSEQAVRDPVLVFRIGAGQVREYTAIL 143
DB 92 SVTVSSKGLTKVY-----HKLGRKAV-IAVSSEQAVRDPVLVFRIGAGQVREYTAIL 143
```

```

Db 676 SVNDVKSLETKTTSVNLNHEIIGNHFINLVDSEKVEQEQPTTQLETDSEFVLPTYQIVE 735
QY 144 D-----PVGYSPKTKSAL-----SDGTHRKRTAPTAEQ-----EN----- 174
Db 736 DSFTESAETNEFSEQOKTLEPTISQOQEVETSESNVPTVEQTKLFHQDNNLFTPLP 795
QY 175 -----QNAKALRKTKDSDANAANKVPAKNGKTHTVRKGETVKQIAAARPKHLTLEQ 226
Db 796 LDLTETIESNALFDSKPKDEKSSDSLEQTF-----KEIKLDS 833
QY 227 VADALLKANGNVSAHGRLAGSVLHIPLNLRKAEQPKQPTA-----KPKAETASM 277
Db 834 TVEV-----FOESSQVATEFTVQPEAVFDEIKTQELQPEATTEWFDHFQDVQPEQT 888
QY 278 PSFSPKQATVEKPKVEKPAKVAAP-----EAKAEKPAVRPEVP 316
Db 889 PQEAKFDPVEIPIQESSQAEFHAQISDEIKLEKTEAVFDHQLENOQSSETVPTPEVT 948
QY 317 AANTAASETAESAQAASAIIDPTDETGNVSEVP-----EQVSABEETESGLFGG 370
Db 949 AFEPETIETQLE-----PSEEDQ-----SEPALQNHPEIVTAEVQ-----IFDG 990
QY 371 SYTLLAGGGAALIALLLRLAQSKEARTESEVPPEEPDLDDAADDGI-----EIT 423
Db 991 TKL-----BDLKEANFDNVNNEVQPKETAEIT 1021
QY 424 F---AEVETPATPEKPNKNDVNDTLALDGESEBELSA-----KOTPDVETDTSNR- 471
Db 1022 FDETKELQETSLEPLSTELKSEAFDNVSEASEAVPEKPKOLETQTEKILBEPKSEP 1081
QY 472 -----IDLDFDSLAANGILSGALTQD-----RETQKRADAMNAIESTDSVYBETFNPY 523
Db 1082 VQQLITEASFDTV--RHEAVFDKNTQQTGLEBEPQVSEAEVVDQTTTDTVGPEA--- 1135
QY 524 NPVEIIVDTPPEPSVAQT-----AENKPEVTDVDFSDNLPNSNNHIGTEETASAKPSPGL 579
Db 1136 -----VFDV--QPKETVEKFDVVENQKVISEPVQEQGGEAVF--EPSAAKFDSP--- 1184
QY 580 AGFLKASSPETILE--KTVAEVQ-----TPELHDFLKVET-----DAVAETAPETPDF 627
Db 1185 VESVQDSQPEFVLEEVQTEPIQVSESPQATEFTVQPEQTPOEAKFDSPVETV-EQPEF 1243
QY 628 N-----AAAD-----LSALLOPAE-----APSV 647
Db 1244 SSEPTQHVSEASFDPEPNYDFDEPNYDFQPSYDGLQSEPVQYDVRPNYDFDEPNYE 1303
QY 648 ENITVETVATPORNATADDLSALLOPSEVPVAAEIVADDLSALLOPAE-APAVREN 706
Db 1304 ---IESKSPQEPQPEVQ-----QGE--AVFEPSEAKFDSPVESVQDSQPEPLLEE- 1352
QY 707 VTETVAETSDFHTAADDLSALLOPAEVP--AVEENVTKTVAETPDNATADDLSALLOPS 764
Db 1353 -VQTEPIQVSESPQATEFTVQPEQTPOEAKFDSPVETIQE-PQVSSPE---VVQPN 1407
QY 765 EYPVAVEENAAETITLTPDSNTSEADALPDLKDGEEETVDSYIYSEEN 813
Db 1408 -----FEERKPEVLEEPQADEIQPEA-----SEBESLDWELLVGNNS 1445

```

RESULT 7

```

CPN_DROME          STANDARD;      PRT;      865 AA.
ID CPN_DROME
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Calphorin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

```

[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphorin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphorin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -I- FUNCTION: Might function as a calcium-sequestering "sponge" to
regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
of Ca(2+) per mole of protein.
CC -I- SUBUNIT: Homodimer (Probable).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
CC -I- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
compound eyes and ocelli.
CC -I- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
development.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L02111; AAA28405.1; -.
DR EMBL; L05080; AAA28420.1; -.
DR PIR; A47282; A47282.
DR PIR; A47283; A47283.
DR FlyBase; FBgn0010218; Cpn.
CC GO: 0005509; F:calcium ion binding; IDA.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match
Best Local Similarity 24.2%; Score 243; DB 1; Length 865;
Matches 172; Conservative 84; Mismatches 285; Indels 170; Gaps 32;

QY 234 ANPNVSAHGRLAGSVLHIPLNLRKAEQPKQPTAKPKAETASMPSEPSKQATVEKPEK 293
Db 60 ATVTTPAPAPIAAASVTPV-----ASVAPPVVAAPTTPPAASPVSTPVAVAQIPVAVSA 112
QY 294 PEAKVAEAEKAEKPAVPEVP-ANNTAASETAAESAPQEAASAIPTPT-----D 344
Db 113 P---VAPPEVAATPTPVQ---IPVAAPVIAATPPVAAASAPTAAVTPVISPVASPPVPA 166
QY 345 ETGNAVSEPVQVSEEEETESGLFGSYTLLAGGGAALIALLLRLAQSKEARTEES 404
Db 167 NTVVPAAPVAAVPAAPVVPAP-----VLAPAVAPAVAPVVAETPAPPVAPVAT 218
QY 405 VPE-EEDLDD-----AADGIEITFAEVEPTPATPEKPNKNDVNDTLALDGESEE 454
Db 219 IPECVAPLIPESVVATKPLAAAEPPVVPATPTFPVAPAAASPHVSVAPAVETAVAP 278

```


Db 404 KPAPAAKAVA---TPKQ-----PAGSGQKPSKADSSSE-----BESSSE 443

Qy 648 ENITETVAETDFNATADDLSALLQPSVPAVENAA--EIVADDLSALLQ-----PAEA 700

Db 444 EEATKKSVTTPKARVTAKAAPSL-----PAKQPRAGGDSSESSSEEEKTPPKP 497

Qy 701 PAVEENVETVAETSDFTAAADDLSALLQPAEVPVAVENVTKTVAETPFDNATADDLSAL 760

Db 498 PAKKKAAGAAVPKPTPVKKAASSESSSSSEDSSEBEK-----KKPKSKATPKPQAG- 550

Qy 761 LQPSVPAVENAA---EITLPTDNTSADALP-DFLKQGEETVDMYSIYSEENIP 815

Db 551 -KANGVPASQNGKAGKSESEEEETQNKAAAGTKPGSGKRKKNETADEAATPQSKVK 609

Qy 816 NNADTSPSSVSGDAPSEAKYDIAEMYLEIGDRDAAEVTKLLLEAEGDVLKRA 871

Db 610 LQTPNTPFKRKGKGRASSPFRVREEIEVDSR--VADNSFDAKRGAGDWGERA 663

RESULT 9

SLPI_CLOTH

ID SLPI_CLOTH STANDARD; PRT: 1664 AA.

AC Q06852;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).

DE OLBP.

GN OLBP.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 10682;

RA MEDLINE=93209931; PubMed=8458832;

RX Fujino T., Beguin P., Aubert J.-P.;

RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CtpA and a protein possibly involved in attachment of the cellulosome to the cell surface.";

RL J. Bacteriol. 175:1891-1899 (1993).

CC -! SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.

CC -! SUBCELLULAR LOCATION: Cell wall.

CC -! SIMILARITY: Contains 4 S-layer homology (SLH) domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X67506; CAA47841.1; --

DR PIR; T18262; T18262.

DR InterPro; IPR008965; Cellul_bind.

DR InterPro; IPR001119; SLH.

DR Pfam; PF00395; SLH, 3.

DR PROSITE; PS01072; SLH DOMAIN; 2.

KW Cell wall; S-layer; Signal; Repeat.

FT SIGNAL 1 28

FT CHAIN 29 1664

FT CELL SURFACE GLYCOPROTEIN 1.

FT REPEAT 36 763

FT REPEAT 36 191

FT REPEAT 207 363

FT REPEAT 409 565

FT REPEAT 607 763

FT DOMAIN 771 1377

FT APPROXIMATE TANDEM REPEATS OF T-P-S-D-E-P.

FT DOMAIN 1378 1449

FT GLY/PRO/SER/THR-RICH.

FT DOMAIN 1453 1494

FT SLH 1 (INCOMPLETE).

FT DOMAIN 1495 1565

FT SLH 2.

FT DOMAIN 1566 1625 SLH 3.

FT DOMAIN 1526 1646 SLH 4 (INCOMPLETE).

SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match 5.4%; Score 239; DB 1; Length 1664;

Best Local Similarity 22.2%; Pred. No. 0.011;

Matches 135; Conservative 78; Mismatches 248; Indels 146; Gaps 28;

Qy 260 AEQPKPTAKPKAETAMPSEPSKQATVEKPV-----EKPEAKVAAPAKAKPAVRPEPV 315

Db 785 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 844

Qy 316 PAANTAASETAASAPQEAASALDTDTGTNAVSEPV---EQVSAEEETESGLFGSY 372

Db 845 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD- 890

Qy 373 TLLAGGGAALIALLLRLAQSKRARRTESVPEEPDLDAAADGGIETFAEVETPA- 431

Db 891 -----EPTSPETPEPIPTDTPSDEPTPSD--EPTPSDEPTPSD 927

Qy 432 --TP--EPAPKNDVNTIALDGESEELSAKQTFDVTETDTPSNRIDLDFSLAAQNGIL 487

Db 928 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD----- 969

Qy 488 SGALTQDEETQKRADADWNAIESTDSVYEPETENPVNVEIVDTPEPE---SVAQTAEN 544

Db 970 -----DEPTP--SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1019

Qy 545 KPETVD--TDFSNLPSNNHIGTEE--TASAKPASPSGLAGFLKASSPETILEKTVAERVQT 601

Db 1020 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1077

Qy 602 PEELHFLKVIYETDAVETAPETPDFAAADLSALLQPAEAPSVEENITETVAETPDFN 661

Db 1078 PSD-----EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1124

Qy 662 ATADDLSALLQSEVPAVENAAEIVADDLSALLQPAEAPAVEENVETVAET-SDFHTA 720

Db 1125 PT-----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1170

Qy 721 ADDLSALLQPAEVPVAVENV-TKTVAEIP-----DFNATADDLSALLQSEVPAVE 770

Db 1171 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD- 1220

Qy 771 ENAAE-ITLETSDNTSEADALPDFLKOGEETVDMYSIYSEENIPNNADT--SFPSES 827

Db 1221 ETPPEPIPTDTPSDEPTPSD-----EPTPSDEPTPSDEPTPSDEPTPSDEPTPEPI 1270

Qy 828 GSDAPSE 834

Db 1271 PTDTPSD 1277

RESULT 10

MINT HUMAN

ID MINT HUMAN STANDARD; PRT: 3664 AA.

AC Q96758; Q9H9A8; Q9NH55; Q9TQ01; Q9Y556;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).

GN MINT OR SHARP OR KIAA0929.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION, RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3; RAR AND MTAIL1.

RC TISSUE=Liver, and Pituitary;

RX MEDLINE=21231190; PubMed=11331609;

RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,

QY 850 DAAAEVQX 858
 Db 1890 DLKLTQVTK 1898

RESULT 12
 IGA4_HAEIN STANDARD; PRT; 1849 AA.
 AC P45386;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
 GN IGA.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NHTI HK61;
 RC MEDLINE=92234949; PubMed=1373717;
 RA Poulsen K., Reinholdt J., Kilian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases.";
 RL J. Bacteriol. 174:2913-2921(1992).
 CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
 CC certain Pro-Xaa bonds in the hinge region. No small molecule
 CC substrates are known.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: Belongs to peptidase family S6.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M87491; AAA24968.1; -
 CC MEROPS; S06.001; -
 CC InterPro; IPR006315; Autotransport.
 CC InterPro; IPR005546; Autotransporter.
 CC InterPro; IPR000710; Peptidase S6.
 CC InterPro; IPR004899; Pectactin.
 CC Pfam; PF03797; Autotransporter; 1.
 CC Pfam; PF02395; IGA1; 1.
 CC Pfam; PF03212; Pectactin; 1.
 CC PRINTS; PR00921; IGASERPTASR.
 CC TIGRfam; TIGR01414; autotrans barl; 1.
 CC Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1021
 FT PROPEP 1022 1849 IMMUNOGLOBULIN A1 PROTEASE.
 FT ACT SITE 299 299 HELPER PEPTIDE (POTENTIAL).
 FT ACT SITE 299 299 PROBABLE.
 SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 5.3%; Score 232.5; DB 1; Length 1849;
 Best Local Similarity 19.2%; Pred. No. 0.023;
 Matches 185; Conservative 140; Mismatches 377; Indels 261; Gaps 41;

QY 35 LKNNRQIKLIARSAVAVASPOAHG-----LGGINI-QSNLDEPFSSGSIIVT 80
 Db 767 LYSGRNVANITSNITASNAQVHIGYKTDGTVCVRSYDTGYVTCHNSLNSEKALNSFNPT 826

QY 81 GERAKA-LLGGGSVTVSEKGLTAKVHKLGKAVIANVSSEQAVRDPVLVFRIGAGQVREY 139
 Db 827 NLRGNVNLTENASFTLIGKANLFGTIQSIGTSQVNLKENS- - - - -WHLTGSNTVQL 878
 QY 140 TAILDPVGYSPKTKSALSDGKTHRTAPTAAESQENONAKALKR- - - - -TDKK 186
 Db 879 N- - - - -LTNGHHILNAQNDANKVTTNTLVNSLSGNGSFYVWDFTNK 923
 QY 187 DSANAAVKPAYNGKTHV- - - - -RKGETVKQIAAAIRPKH- - - - -LTLRQVADA- - - - -LLKANP 236
 Db 924 SNKVVVWKSATGNFTLQVADKTGE- - - - -PNHNLTLFDASNATNNLEVTLANG 973
 QY 237 NV- - - - -SAHGRL- - - - -PAGSVLHPNLNRKAKBQPKQAK- - - - -PKA 272
 Db 974 SVDRGAWKYKLRNVNGRYDLYNPEVKRNTQVDTTITTPNDIQADAPSQSNNEETAR 1033
 QY 273 ETASMPSEPSKQATV- - - - -EKPVEKPEAKVAPEAKAEKPAVRPVPVPAANTAA- - - - -322
 Db 1034 ETPVPPAPATESAIASEQPETEP- - - - -AETAQPAEMEETNTANSTAPKSDTATQTEPNSE 1092
 QY 323 - - - - -SET- - - - -AAESAPQEAASAIIDPTTDTETGNVASEPVEQS 358
 Db 1093 SVSEETTEKVAENPPQENETVAKNEQEATEPTPQNGEVAKEQPTVEANTQTNEATQSEG 1152
 QY 359 ABEETESGLFGGSYTLILAGGGNAALIALLLLLRLAQSRRTEESVPEEPD- - - - -LDDAA 416
 Db 1153 KTEETV- - - - -AETKSEPTSVTSVENQPKTVSQST 1185
 QY 417 DGIEI- - - - -TFAEVET- - - - -PATPEAPKNDVNDTLALDGESEELSAKQ 459
 Db 1186 EDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAEPAPEVPTDTHAEAAQLCQQTPT 1245
 QY 460 TFDVETDTPSNRIDLDFDLAAQNG- - - - -ILSGALTQDEETQKRADAMNAIESTDS 513
 Db 1246 VAAAEETSPNSKPAEETQOPSEKTAEPVTPVVSSENTATQPTETEETAKVE- - - - -KEKTOE 1302
 QY 514 VVEPETFPYNPVEIVIDTPPEPSVAQTAENKEVTDTPSDNLPNSNHIGTEETASAKP 573
 Db 1303 V- - - - -PQVASQESPKQ- - - - -EQPAAKPQAT- - - - -KPAEPA- - - - -RENVLTTKNVG-EPQPAQP 1351
 QY 574 ASPS- - - - -GLAGFLKASSPETILEKTVAEVQTPPEELHDFLKVYETAETAETPETPDFAA 630
 Db 1352 QTQSTAVPTTGTAAANSKPAAPQAKQKQTPPEARENSTVNTKEPQSOISATVSTEQPA 1411
 QY 631 ADDLSALLQPAEAPSVEENITETVAETPDNFATADLSALLQPSVPAVEENAAEIVADD 690
 Db 1412 KETSSNVEQAPENSINTGSATTMTET- - - - -AEKSKPQME- - - - -TVTEND 1454
 QY 691 LSALLQPAEAPAVEENVTETVAETSDFTAAADDLSALLQPAEVPVAVEENVTVAEIPDP 750
 Db 1455 - - - - -RQP- - - - -EANTVADNSVANNSESE- - - - -SKSRRRSVSPQKETSABETTASTQ- - - - -1501
 QY 751 NATADDLSALLQPS- - - - -VPAVEENAAEITLTPDSNTSEADALPDFLKDGEETVDWSEI 807
 Db 1502 ETTVDNSVSTPKPSRRRTSRVQTSYE- - - - -PVLPTEENAEENVO- - - - -1545
 QY 808 YLSENIIPNNAD- - - - -TSFSESVGSAPSEAKYDLAEWYLEIGDRDAAAEVOKLLEE 862
 Db 1546 - - - - -SGNNVANSOPALNLTSKNTNVLNMAKAQF- - - - -VALNVG- - - - -KAVSQHISOLENN 1597
 QY 863 AEG 865
 Db 1598 NEG 1600

RESULT 13
 HMW1_MYCPN
 ID HMW1_MYCPN STANDARD; PRT; 1018 AA.
 AC Q50365; Q50348; Q50349; Q9R5R4;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

GN NEPH OR NPH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143506;
RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites.";
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT in situ detection.";
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN [3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderton B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H).";
RL FEBS Lett. 209:203-205(1986).
RN [4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -!- FUNCTION: Neurofilaments usually contain three intermediate
CC filament proteins: L, M, and H which are involved in the
CC maintenance of neuronal caliber. NF-H has an important function in
CC mature axons that is not subserved by the two smaller NF proteins.
CC -!- PTM: There are a number of repeats of the tripeptide K-S-P, NPH is
CC phosphorylated on a number of the serines in this motif. It is
CC thought that phosphorylation of NPH results in the formation of
CC intermediate cross bridges that are important in the maintenance
CC of axonal caliber.
CC -!- PTM: Phosphorylation seems to play a major role in the functioning
CC of the larger neurofilament polypeptides (NF-M and NF-H), the
CC levels of phosphorylation being altered developmentally and
CC coincident with a change in the neurofilament function.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 783.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M37227; AAA41693.1; ALT_FRAME.
CC EMBL; X13804; CAA32038.1; ALT_FRAME.
CC EMBL; M21964; AAA41695.1; -
CC EMBL; J04517; AAA41692.1; -
CC PIR; S02003; S02003.
CC InterPro; IPR001664; IP.
CC Pfam; PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
CC NON_TER 1 1

FT	DOMAIN	276	641	51 X 3 AA TANDEM REPEATS OF K-S-P.
FT	CONFLICT	164		L -> I (IN REF. 2).
FT	CONFLICT	185		I -> S (IN REF. 2).
FT	CONFLICT	193		L -> T (IN REF. 2).
FT	CONFLICT	199		M -> T (IN REF. 2).
FT	CONFLICT	346		K -> N (IN REF. 1).
FT	CONFLICT	373		A -> V (IN REF. 1 AND 4).
FT	CONFLICT	482		G -> E (IN REF. 2 AND 4).
FT	CONFLICT	485		P -> S (IN REF. 2).
FT	CONFLICT	570		RK -> KE (IN REF. 2 AND 4).
FT	CONFLICT	591		P -> T (IN REF. 2 AND 4).
FT	CONFLICT	727		A -> V (IN REF. 4).
FT	CONFLICT	757		AAP -> GST (IN REF. 4).
FT	CONFLICT	769		T -> L (IN REF. 2).
FT	CONFLICT	775		R -> P (IN REF. 2 AND 4).
SQ	SEQUENCE	831 AA;	89486 MW;	1B0973C3F13BF768 CRC64;
Query Match 5.0%; Score 220.5; DB 1; Length 831;				
Best Local Similarity 21.2%; Pred. No. 0.027;				
Matches 182; Conservative 120; Mismatches 344; Indels 211; Gaps 40;				
QY	50	VAASFOAHAGLGLNIQSLD-EPPSGSITVTGEEAKALLGGSVTVSEKGLTAKVHKLG	108	
DB	145	MAAQRLREYQDL--LNVKALDLLEIAAYKLEGEBCR--IGFGPIPF--LTEGLPKIP	197	
QY	109	DKAV-IAVSSQAV-----RDPVLVPRIGAGQV-REYTAIILDPVGYSPKTKSALSDG	159	
DB	198	SMSTHIKVKSEKIKVSEKETVIVEEQTEETQVTEETEEDEKAQGESEAEERGG	257	
QY	160	KTHRTAPTAE--SOENQNAKALRTDKDSANAAVKPAYNGKTHVTKGVTKQIAAAI	217	
DB	258	EEAATTSPPAEFAASEKETKSPVKEEAKSPAEK-SPA-----EAKSPAEAK	304	
QY	218	RPKHLTLQVADALIKANPNVSAHRLRAGSVLYHLPNLRKAEQKPKQAKPKAETASM	277	
DB	305	SP-----AEVSPAVAKSPAEVKSPAEAKSPAEVKSATVKSPEG-AKS	359	
QY	278	PSEPSQATVKKPV-----KPEAKVAAP-EAKAEKPAVPEVPAANTAAETAES	329	
DB	360	PAEAKSPAEVKSPAEAKSPAEAKSPAEVKSPEAKSPAEVKSATVKSPEAKS	419	
QY	330	APQEAASAAIDTPTDETGNVSEPV-----QVSAEETESGLFGSGYTLILAGGAALI	384	
DB	420	PAEVKSPVTVKSPAE-----AKSPVEVKSPAVKSPEAKSP-----	456	
QY	385	ALLLLRLAQSKRRARTEESVPEEPDLDDAADDGIEITFAEVETPATPEPAKNDVNT	444	
DB	457	-----AGAKSPAEKSPVAKSP-----AEAKSPAGAKPAE-----	488	
QY	445	LALDGESEELSAKOTFDVETDTPSN-RIDLDLSLAQNGILSGA--LTODEETQKA	501	
DB	489	--AKSPAEAKSPAEAKSPAEAKSPAEVKSPEAKSPVKSPEAKSLAEAKSPEAK	546	
QY	502	DADWNAIETSDSVSEPTFNPVNPVIEVIDTPEPESVAQTAENKPKETVDTDFSDNLPNN	561	
DB	547	SPVKGEIKPPAEVKPE--KAKSPMKKEAKSP-----KAKTLDV-----	584	
QY	562	HIGTSETASAKSPASGLAGFLKASSPETILKTVAEVQTPPELHDFLKVYETDAVETA	621	
DB	585	-----KSPAEKPAKEEAKRPADINSPEQVKSPEAKSPEK-----EETRTKVA	631	
QY	622	PTTPDFAAADLSALLQPAEAPSVEENTVETPDPFNATADLSALLQSEVPAAVE	681	
DB	632	PKKEEVKSPVEEVKAKEPPKK--VVEEKTPTATPKTEVKESKDEAP---KGAQKPAEE	685	
QY	682	NAAEIVADDLSALLQPAEAP--AVENVTETVAETSDFTTAADDLSALLQPAEVPAAVEEN	739	
DB	686	K-----EPLTE--KPKDSFGAEKKEAKEKAAAPPEET-----PAKLVKKEA	727	
QY	740	VTKTVAEIPDFNATADLSALLQPS-----EVPAVEENAABITLETTPSDNTSADAL	791	
DB	728	KPKKEAE---DAKAKEPS---KPSSEKPKPKKEVPAAPPEK-----DTKEETTESK-	774	

QY	792	PDFLKGEEETVDSIYLSEENIPNNADTSPFSESVGSDAFSEAKYDLAEMYLEIGDRDA	851
Db	775	-----REEKPKMEAKAKEE-----DKGLPQE-----PSKPKTEKAERSSSTDQKD-	814
QY	852	AAETVQKLLLE--AEGD	866
Db	815	-SQPSEKAPEDKAAKGD	830

Search completed: September 2, 2004, 19:18:17
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2004, 19:14:12 ; Search time 139 Seconds
(without alignments)
1997.525 Million cell updates/sec

Title: US-09-743-674-2

Perfect score: 4404

Sequence: 1 MPAGRLPRRCPPMNTKFTDCT.....EEAEGDVLKRAQALQALGII 880

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4404	100.0	880	16 Q9JST3	Q9JST3 neisseria m
2	4273.5	97.0	875	16 Q9K147	Q9K147 neisseria m
3	4270.5	97.0	875	2 O86394	O86394 neisseria m
4	436	9.9	962	16 Q8XX7	Q8XXX7 ralistonia s
5	387.5	8.8	947	16 Q87Y10	Q87Y10 pseudomonas
6	381.5	8.7	919	16 Q9HZA6	Q9HZA6 pseudomonas
7	353.5	8.0	1621	16 Q9KTA5	Q9KTA5 vibrio chol
8	347	7.9	1351	16 Q8DB34	Q8DB34 vibrio vuln
9	340	7.7	911	16 Q88LE1	Q88LE1 pseudomonas
10	334.5	7.6	927	2 O87015	O87015 pseudomonas
11	321.5	7.3	717	16 Q886D3	Q886D3 pseudomonas
12	319.5	7.3	2768	5 Q9VC00	Q9VC00 drosophila
13	316	7.2	1822	2 Q07290	Q07290 streptococ
14	315.5	7.2	1786	5 Q9U0P0	Q9U0P0 plasmodium
15	312.5	7.1	673	16 Q8XTC4	Q8XTC4 ralistonia s
16	312	7.1	801	5 Q23635	Q23635 caenorhabdi

ALIGNMENTS

RESULT 1

Q9JST3 ID Q9JST3 PRELIMINARY; PRT; 880 AA.

AC Q9JST3; DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Neisseria-specific antigen protein, Tspa.

GN TSPA OR NMA2146.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=65699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=22491 / Serogroup A / Serotype 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Farhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;

RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.";

RL Nature 404:502-506(2000).

DR EMBL; AL162758; CAB85358.1; --

DR PIR; G81786; G81786.

DR GO; GO:0015998; P:cell wall catabolism; IEA.

DR InterPro; IPR002482; LysM.

DR InterPro; IPR001440; TPR.

DR Pfam; PF01476; LysM; 1.

KW Complete proteome.

SQ SEQUENCE 880 AA; 93153 MW; A1701AFFE849338A CRC64;

Query Match 100.0%; Score 4404; DB 16; Length 880;
Best Local Similarity 100.0%; Pred. No. 5.5e-205;
Matches 880; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
Db 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
QY 61 GGLNIQSNLDPPFGSSITVTGEEAKALLGGGSVTVSEKGLTAKVHKLGDKAVIIVSSBQA 120
Db 61 GGLNIQSNLDPPFGSSITVTGEEAKALLGGGSVTVSEKGLTAKVHKLGDKAVIIVSSBQA 120
QY 121 VRDVLVFRIGAGAVREYTAILOPVGVSPKTSALSDGKTHRTKAPTAESEONAKAL 180
Db 121 VRDVLVFRIGAGAVREYTAILOPVGVSPKTSALSDGKTHRTKAPTAESEONAKAL 180
QY 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRIRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
Db 241 HGRIRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
QY 301 PEAKAEKPAVRPEPVPAAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAA 360
Db 301 PEAKAEKPAVRPEPVPAAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAA 360
QY 361 EETESGLFGGSYTLGAGGAAIALIALLRLAQSRRARRTEESVPEEPDLDAAADGI 420
Db 361 EETESGLFGGSYTLGAGGAAIALIALLRLAQSRRARRTEESVPEEPDLDAAADGI 420
QY 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTPSNRIDLDPSLA 480
Db 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTPSNRIDLDPSLA 480
QY 481 AAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFPNPNVEIVIDTPEPESVAQ 540
Db 481 AAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFPNPNVEIVIDTPEPESVAQ 540
QY 541 TAENKPEVTVDITDPSNLPSNNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAEQ 600
Db 541 TAENKPEVTVDITDPSNLPSNNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAEQ 600
QY 601 TPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAETPDF 660
Db 601 TPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAETPDF 660
QY 661 NATADDLSALLQPSSEVPAVEENAAEIVADDLSALLQPAEAPAEVENVETVAETSDPHTA 720
Db 661 NATADDLSALLQPSSEVPAVEENAAEIVADDLSALLQPAEAPAEVENVETVAETSDPHTA 720
QY 721 ADDLSALLQPAEVPVEENVTKTVAEIPDFNATADDLSALLQPSSEVPAVEENAAEITLET 780
Db 721 ADDLSALLQPAEVPVEENVTKTVAEIPDFNATADDLSALLQPSSEVPAVEENAAEITLET 780
QY 781 PDSNTSEADALPDFLKDGEEETVDSIYI SEENIPNNADTSPPSSVSGDAPSEAKYDLA 840
Db 781 PDSNTSEADALPDFLKDGEEETVDSIYI SEENIPNNADTSPPSSVSGDAPSEAKYDLA 840
QY 841 EMYLEIGDRDAAAETVQKLLLEAEGDVLKRAQALAEELGI 880
Db 841 EMYLEIGDRDAAAETVQKLLLEAEGDVLKRAQALAEELGI 880
```

RESULT 2

```
Q9K147
ID Q9K147 PRELIMINARY; PRT; 875 AA.
AC Q9K147;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TspA protein.
GN NMB0341.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
```

```
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Pedon J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterio H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.; B strain
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815 (2000).
DR EMBL; AE002391; AAF40784.1; --
DR PIR; C81209; C81209.
DR TIGR; NMB0341; --
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01476; LysM; 1.
KW Complete proteome.
SQ SEQUENCE 875 AA; 92488 MW; 1F921520C167D090 CRC64;
```

Query Match 97.0%; Score 4273.5; DB 16; Length 875;
Best Local Similarity 97.1%; Pred. No. 1.1e-198;
Matches 858; Conservative 5; Mismatches 8; Indels 13; Gaps 2;

```
QY 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
Db 1 MPAGRLPRCPMTKFTDCTRSNRIQPTTHRGYILKNNRQIKLIAASVAAASFOAHAGL 60
QY 61 GGLNIQSNLDPPFGSSITVTGEEAKALLGGGSVTVSEKGLTAKVHKLGDKAVIIVSSBQA 120
Db 61 GGLNIQSNLDPPFGSSITVTGEEAKALLGGGSVTVSEKGLTAKVHKLGDKAVIIVSSBQA 120
QY 121 VRDVLVFRIGAGAVREYTAILOPVGVSPKTSALSDGKTHRTKAPTAESEONAKAL 180
Db 121 VRDVLVFRIGAGAVREYTAILOPVGVSPKTSALSDGKTHRTKAPTAESEONAKAL 180
QY 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTOKOSANAAPKAYNGKTHTVRKGETVQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGRIRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
Db 241 HGRIRAGSVLHLPNLNRIKAEQPKQTAETASMPSEPSKQATVEKPEKPEAKVAA 300
QY 301 PEAKAEKPAVRPEPVPAAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAA 360
Db 301 PEAKAEKPAVRPEPVPAAANTAAETAESAPOEAAASALDPTDGTGNVSEPEVQVSAA 360
QY 361 EETESGLFGGSYTLGAGGAAIALIALLRLAQSRRARRTEESVPEEPDLDAAADGI 420
Db 361 EETESGLFGGSYTLGAGGAAIALIALLRLAQSRRARRTEESVPEEPDLDAAADGI 420
QY 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTPSNRIDLDPSLA 480
Db 421 EITFAEVTPTATPEPAPKNDVNDTLALDGESEELSAKQTFDVTDTPSNRIDLDPSLA 480
QY 477 DSLAAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFPNPNVEIVIDTPEPE 536
Db 477 DSLAAQNGILSGALTQDEETOKRADADWNAIESTDSVVEPETFPNPNVEIVIDTPEPE 536
QY 537 SVAAQTAEKPEVTVDITDPSNLPSNNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAE 596
Db 537 SVAAQTAEKPEVTVDITDPSNLPSNNHIGTEETASAKPSPGLAGFLKASSPETILEKTVAE 596
QY 597 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAE 656
Db 597 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAE 656
QY 601 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAE 660
Db 601 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADLALLOPAEAPSVEENITETVAE 660
```

```

QY 657 TPDFNATADLSALLQSPSEVPVAVENAAEIVADDSALLQPAEAPAVEENVETVETVTS 716
Db 661 TPDFNATADLSALLQSPSEVPVAVENAAEIVADDSALLQPAEAPAVEENVETVETVTS 720
QY 717 FHTAADDLSALLQPAEAPAVEENVTKVVAEIPDFNATADLSALLQSPSEVPVAVENAAE 776
Db 721 FNATADLSALLQPAEAPAVEENAAETV-----ADDSALLQPAEAPAVEENAAE 771
QY 777 TLETPDSNTSEADALPDFLKGSEETVDSIYILSEENIPNNADTSFSESVDGSDAPSEAK 836
Db 772 TLETPDSNTSEADALPDFLKGSEETVDSIYILSEENIPNNADTSFSESVDGSDAPSEAK 831
QY 837 YDLAEMYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEGLGI 880
Db 832 YDLAEMYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEGLGI 875

RESULT 3
O86394
ID O86394 PRELIMINARY; PRT; 875 AA.
AC O86394;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TspA protein.
GN TSPA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=B:15;PI.16;
RX MEDLINE=99307215; PubMed=10377136;
RA Kizil G., Todd I., Atta M., Borriello S.P., Ait-Tahar K.,
RT "Identification and characterisation of TspA, a major CD4+ T-cell and
RL B-cell stimulating Neisseria-specific antigen.";
DR EMBL; AJ010113; CAA09002.2; -
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR SQU SEQUENCE 875 AA; 92548 MW; 1F920E217A677091 CRC64;

Query Match
Best Local Similarity 97.0%; Score 4270.5; DB 2; Length 875;
Matches 858; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 MPAGRLPRRCPMWTKFTDCTRSNRIQPPTHRGVYILKNNRQIKLIAASVAVASFOAHAGL 60
Db 1 MPAGRLPRRCPMWTKFTDCTRSNRIQPPTHRGVYILKNNRQIKLIAASVAVASFOAHAGL 60
QY 61 GGLINTQSNLDEPFSGSITVTGEAKALGGGVTVSEKGLTAKVHKLGDKAVIAYSSEA 120
Db 61 GGLINTQSNLDEPFSGSITVTGEAKALGGGVTVSEKGLTAKVHKLGDKAVIAYSSEA 120
QY 121 VRDPLVFRIGAGAVREYATLDPVGYSPKTSALSDGKTHRKTAFTAESQENAKAL 180
Db 121 VRDPLVFRIGAGAVREYATLDPVGYSPKTSALSDGKTHRKTAFTAESQENAKAL 180
QY 181 RKTDKKSANAAVKPAYNGKTHVTKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
Db 181 RKTDKKSANAAVKPAYNGKTHVTKGETVKQIAAAIRPKHLTLEQVADALLKANPNVSA 240
QY 241 HGLRAGSVLHPIPNLRKAEQPKQTAKPKAETASMPSEFSQATVEKPKPEAKVAA 300
Db 241 HGLRAGSVLHPIPNLRKAEQPKQTAKPKAETASMPSEFSQATVEKPKPEAKVAA 300
QY 301 PEAKAEKPAVRPVPVPAANTAASETAAESAQAQAAASAITPTDGTGNVSEPVQVSAE 360

```

RESULT 4

```

Q8XXX7
ID Q8XXX7 PRELIMINARY; PRT; 962 AA.
AC Q8XXX7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable transmembrane protein.
GN RSC1986 OR RS03415.
OS Ralstonia solanacearum (pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Chandler M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Gaspin C., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Siguer P., Thebault P., Moisan A., Robert C., Saurin W., Schiek T.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646067; CAD15688.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR001854; Ribosomal_L29.
DR InterPro; IPR001440; TPR.

```

```

Db 301 PEAKAEKPAVRPVPVPAANTAASETAAESAQAQAAASAITPTDGTGNVSEPVQVSAE 360
QY 361 EETESGIFDGLFGSGSYLLLAGGGAALILLRLAQSKRARTESVPEEPDLDDAA 416
Db 361 EETESGIFDGLFGSGSYLLLAGGGAALILLRLAQSKRARTESVPEEPDLDDAA 420
QY 417 DDGIEITFAVEVETPATPEPAKNDVNDTLADGSESEELSAAKOTFDVETTPSNRIDLF 476
Db 421 DDGIEITFAVEVETPATPEPAKNDVNDTLADGSESEELSAAKOTFDVETTPSNRIDLF 480
QY 477 DSLAAQNGILSGALTQDEETQKRDADADWNAIESTDSVYEPETFPNPNVPIVIDTPEPE 536
Db 481 DSLAAQNGILSGALTQDEETQKRDADADWNAIESTDSVYEPETFPNPNVPIVIDTPEPE 540
QY 537 SVAQTAEKPKETVDTDFSNLPSNNHIGTEETASAKPASGLAGFLKASSPETILEKT 596
Db 541 SVAQTAEKPKETVDTDFSNLPSNNHIGTEETASAKPASGLAGFLKASSPETILEKT 600
QY 597 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADDSALLQPAEAPSEVENTITVAE 656
Db 601 AEVQTPPEELHDFLKVYETDAVETAPETPDFNAAADDSALLQPAEAPSEVENTITVAE 660
QY 657 TPDFNATADLSALLQSPSEVPVAVENAAEIVADDSALLQPAEAPAVEENVETVETVTS 716
Db 661 TPDFNATADLSALLQSPSEVPVAVENAAEIVADDSALLQPAEAPAVEENVETVETVTS 720
QY 717 FHTAADDLSALLQPAEAPAVEENVTKVVAEIPDFNATADLSALLQSPSEVPVAVENAAE 776
Db 721 FNATADLSALLQPAEAPAVEENAAETV-----ADDSALLQPAEAPAVEENAAE 771
QY 777 TLETPDSNTSEADALPDFLKGSEETVDSIYILSEENIPNNADTSFSESVDGSDAPSEAK 836
Db 772 TLETPDSNTSEADALPDFLKGSEETVDSIYILSEENIPNNADTSFSESVDGSDAPSEAK 831
QY 837 YDLAEMYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEGLGI 880
Db 832 YDLAEMYLEIGDRDAAEIVQKLEEAEGDVLKRAQALAEGLGI 875

```

DR InterPro: IPR008941; TPR-like.
 DR Pfam: PF01476; IYEM: 1.
 DR PROSITE: PS00579; RIBOSOMAL_L29; 1.
 KW Complete proteome.
 SQ SEQUENCE 962 AA; 99116 MW; EFDF380852P0730 CRC64;

Query Match 9.9%; Score 436; DB 16; Length 962;
 Best Local Similarity 22.8%; Pred. No. 5.2e-13;
 Matches 238; Conservative 147; Mismatches 351; Indels 306; Gaps 41;

QY 46 ASVAVASP-----QAH-AGLGINTQSNLDEPFGSITVTG---EAKALLG--GGSV 93
 DB 19 SAVAIASAALLIQPAHAAGFGALHVRSSLGQPLQAEIDLGSVTEEEAQNVLAKLASPD 78

QY 94 TVSEKGLT-----AKVHKLGDKAVIASVSEQVRDP---VLFRIGAGAV--REY 139
 DB 79 AYQAGUTYNPITVTLASLVRQGGSYVVRVISAQPVAEFPVILVDLITWASGRVSEAY 138

QY 140 TAILDPVGYSPKTKS-----ALSDGKTHRKTAFTAESQE-----NONAKALRKTDKK 186
 DB 139 TFLDPAGSSNTPNFAPTPPVQATTGCVVDSTPAPVAAAPQAPAAAPRAAPARQAAR 198

QY 187 DSANAAPKAPYNGKTHVTKETVKQIAA-AIRPKH-LTLEQVADALLKANPNVSAHG-- 242
 DB 199 PQADAAAAPSSGAGYTVQRGDSLYDVASNAVQGDGVSQDMLLALYRNNPKAFIGNI 258

QY 243 -RLRAGSVLHLPN---LNRIKAEQPKQ-----TAKPKA 272
 DB 259 NLRGTGSLVTFPSAAEAKQVSRREARVVAQTSGFAGYRSLRATAEANAATDTSARQ 318

QY 273 ETASNPSPSPKQAT-----VEKPVKEPKAVA 299
 DB 319 QSGSVARVQOQATPSASERDELRLSKADRTGKAAATAGARAEELVAKELKEMESRVA 378

QY 300 APE-----AKAEKAVRPEVP-----AANTAASETAESAPO 332
 DB 379 QLEKNLSEMHQHLIEVKNAELAKAAAKPAGAAPSPATAAPAAVTAANAPASASATPA 438

QY 333 E---AASAIOT-----PTDETGNVSEVPQVSA----- 359
 DB 439 QAPAAAASAEFASATATAGASAPAAATPAASAPVAEASAPAAAPKAPVVAPOPPA 498

QY 360 -EETESGIFGYSYTIILAGGAALTALLLRLAQSK--RARRTESVPEEPELDLDA 416
 DB 499 EESFESSLLGNPMALGLGLVALAGLVYRRQKPEQAHQFQDSLLSQESTWMAGA 558

QY 417 DD-----GLEITFAVETATPEPAPKNDVNTLALDGSSELSAK 458
 DB 559 NSLFGAAGQSDITSQHSYFGADFRIGGNGESNEVDPTAEADVITYAGRDVQAEELL--R 616

QY 459 QTFDVEDTTPSNRIDLDPDSLAAQNGILSGALTQDEETQKR---ADADWNAIESDTSVY 515
 DB 617 EALEQHPEQAILRLK-MEIVANRQDA--HGQTTAEEMLAQVGAASPDW-----AEAAA 668

QY 516 EPETENPNVNFIVIDTPPEPESVAQTAENKPTVDTDFSDNLPSSNNHIOTETASAKPAS 575
 DB 669 LGRKTPDANPLYLTQV-----GDGHEHQVAAD-----DRHGHAQ-----GAVAAA 708

QY 576 PSGLAGFLKASSPETILEKTVAEVQT--PEELHDLFLKVYETDAVATAPETPDFN----- 628
 DB 709 GAALAGMGAAAAAFAKFTVTGTTTRRGEHWTTVDPMGDFMSPSTKAPQLADLEPLES 768

QY 629 ----AAADLSLALQAPAPSEENITETVA-----ETP--DFNATADDLSALLQSPSEVA 678
 DB 769 FPAPAGEFITAPLOAE--VFQPLAEPEAPPALPQTDPAFHAGE---AFQPASVPP 822

QY 679 VERNAAEIVADDLSLALQAPAPAEVENTETVAETSDHEHTAADDLSLALQAPAEVAVEE 738
 DB 823 LHMDLSDL-SDLNTPAPVAETAV-----EPAPAAVAADSLFAPAAQPADLP----- 868

QY 739 NVTKTVAEI PDFNATADDLSLALQSPSEVPAVEENAAEI-TLEPTDNGNTSEADALPDFLKD 797
 DB 869 -----ETPVQLDAAPQPEVVAQQEPLVTMRLDNLPHLSAEKGIDCVRD 915

QY 798 GBEETVDMYSIYLSEENIPNNADTSFSPSEVSGSDAPSEAKYDLAEWYLEIGRDAAAATVQ 857
 DB 916 -----LQIKFDLAKAYIEIGDKEGARELLQ 940

QY 858 KLEEEARGDVLKRAQALAAQELG 879
 DB 941 EVLDLGDPSFHAERQAALMRQIG 962

RESULT 5
 Q87Y10 PRELIMINARY; PRT; 947 AA.

ID Q87Y10
 AC Q87Y10;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN PSTO3818.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 EX NCBI_TaxID=323;
 EN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
 RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
 RA White O., Fraser C., Collner A.;
 RA "Complete sequence of Pseudomonas syringae";
 RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE016869; AA057286.1; -
 DR TIGR; PSPTO3818; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR008941; TPR-like.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 947 AA; 99429 MW; 84789769FCF9B69 CRC64;

Query Match 8.8%; Score 387.5; DB 16; Length 947;
 Best Local Similarity 22.6%; Pred. No. 1.1e-10;
 Matches 230; Conservative 133; Mismatches 377; Indels 279; Gaps 39;

QY 44 IAAVAVAAASFOAHAGLGLNLIQSNLDEPFGSITVTGEEAKALLGGSVTVSEKGLTAK 103
 DB 11 IAAASALSSGMAQALGLGELSVKSTLNQPLVAEILT--DAQGL--NAAQVVPFLATTAD 66

QY 104 VHKLK-----DKAVIASVSEQVRDPVLVFRIGA---GAQVREYT 140
 DB 67 FAQLGVSRQELLNDLITFTFPIDPNGESVLRTTSKPVVPYVKFLVQLWPNGLVRDYS 126

QY 141 AILDVPVGYSPKTKSALSDGKTHRKTAFTAESQ-----ENQNAKALRKTOKDSANAAVKP 195
 DB 127 LLLDPPKYSPEAAAAAAPAPAPAAAPAAQATVAPSTEAAPAPAPAPAPAPAPAPAP 186

QY 196 ---AYNGKTHVTKETVKQIAAATRPKHLTLEQVADALLKANPNVSAHG---RLRAGSV 249
 DB 187 APAADKQAVYVANNITLWEIAAKVRTGG-TVQQTMLAQNLNPDFAFMGNINRLKKGQV 245

QY 250 LHIPNLNRIKAEQPKQ-----TAKPAETASMPSEPS 282
 DB 246 LRLPTQQTTA-LPQQAIAEVSROYSAWKEGRRLPTGTQVDAITRRDRAGAPSKVDTS 304

QY 283 KQATVEKPVKEPEAK----- 297
 DB 305 DNLSLVSAAGKPAKAGGAGDADLGNKLAVAQEAQLDTRDRNDELKSRMNDLQSLDKLQR 364

```

QY 298 -----VAAPEAKAEKPAVRPEPVPAAANTAASATAESAPQOE-- 333
Db 365 LIELKNGQLAKMQAAGAAVPPVAATPEATTPANAAVPAASLVANGAPVKPAGEIAPEDAL 424
QY 334 -AASALDITPTDEGNNAVSEVQVSAEETESGLFGSYTLLAGGGAALIALLLLR 392
Db 425 PAGAQAQVATPAADQPLAV-BFVAATEDDDMLQKAL-DNPITLLGLIGGAALLIALLLFL 482
QY 393 AQSKRRAR-----TEES--VPBEPDLDDAADGGIETTFAEVETPA----- 431
Db 483 ARRAAKAEAEKHKMARALAESDFVSDMDMDAPQASFGDLVPPNVRMGAAAGAA 542
QY 432 -----TPEAPKXNDVNTLALDGESEELSAKQTFDVTTPSNRIDLDFSLAAQ-NG 485
Db 543 AABERSADPLVQAEIHIAVGRMNAQAVELLEBA-----VKDDPARDIRLKLMEIVAEQGN 598
QY 486 ILSGALTODEETQKADADNATSTDSVTEPEFNFYNEVEIVDTPEPESVAQTAENK 545
Db 599 KAYAHERKLVAAKGAEEVEQLEERHS-----TLKFPVAPVVLFPVAPQASEPAPASAA 653
QY 546 PETVDTDFDNLPSNNHIGTEASAKPASPSGLAGFLKASSPETILEKTVAEVQTPPEL 605
Db 654 PAV-----AAAVAASAALAEALDAKYVEELLADDSAE-QPPEPV 693
QY 606 HDFLKY-----ETDAVAETAPETP-----DFNAAADDLSALLQPAEAPSVEENITETVAETP 658
Db 694 AEPEPVVPEPELAAAPVEEDPFDFDLSLDEFTTTP--QVSTVDNLDLMLDEP 751
QY 659 DFNATADD--LSALLQSEVPAVEENAA--EIVAD--DLALLQPAEAPAVEE--- 705
Db 752 ALSAVDDPELSPESIMQOE-----BARAATTFEDLADFDLDS-----EEDPALKNEDD 802
QY 706 -----NVTETVAETSDFTHTAADDLSALLQPAEVPVAVENKIVA-----EIPDFENATA 754
Db 803 FLILGLSGEPLDGETKVPPISSDLE-LPEDFDSLADETDQASQAFATEIDVNA-- 859
QY 755 DLSALLQSEVPAVEENAAITLETDPDNTSADALPDFLKDGEETVDMWSIYLSEENI 814
Db 860 -ELDRLAQNLEHPDLE-----PRFTAEDAAALDD----- 888
QY 815 PNNADTSPFSESVCSPASEAKYDLAEMYLEIGDRDAAAEVTKLLEBAGDVLKRAQA 873
Db 889 --EPDPDF--MAGTD-EAATKLDLARAYDMGDADGARDILDEV--TEGDDGQKSEA 939

RESULT 6
QHZA6 PRELIMINARY; PRT; 919 AA.
AC Q9HZA6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PA3115.
GN PA3115.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004735; AAG06503.1; -.
DR F01; F83257; F83257.

```

```

DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR000408; Reg_chrom_condens.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF01476; LysM; 2.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS00626; RCC1_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 919 AA; 96928 MW; CE745BCFAD414A CRC64;

Query Match      8.7%; Score 381.5; DB 16; Length 919;
Best Local Similarity 23.0%; Pred. No. 2.1e-10;
Matches 232; Conservative 134; Mismatches 362; Indels 281; Gaps 43;

QY 41 IKLIAASVAVAAAFQAH-GLGINTQSNLDEPFGSGITVTGEEKALLGGGVTVS--- 96
Db 8 VRAIAAA-SVLTSGMAHGLGELITLKSALNQPLDAEIELL--EVDR-LGSEVIPSLAS 63
QY 97 -EKGLTAKVHKL-----GDKAVIASVSEQAVRDPVLPFRIGA---GAQVR 137
Db 64 PEFSKAGVDRLYLTLKFTPVKPKNGKSVIRTSKPVQEPFLNPLVQVLPNGRLLR 123
QY 138 EYTAIILDPVGYSPKTSALSADGKTHRTAPTAEBSQENQNAKALRKTDKDSANAAYK-- 194
Db 124 EYTVLLDPPLYSPQAAASAPQAPV--SAPRATG-----APRAPQAPVETTA 169
QY 195 PAYNGKTHVTKGETVKQIAAIRPKHLILEQVADALLKANPNVSAHG---RLRAGSVLH 251
Db 170 PAGESDYRTV-SNDTLWEIAQRNRTDRVSPQAMLAFOELNPGAFVQDGNRLKSGQVLR 228
QY 252 IPNLNRIKAEQPKFQTKPKAETAS-----MPSEPSK-QAT 286
Db 229 IPTQQMLERSPREALSQVQAQNSQWRGNPAAGSAGARQLDATORNAAGSAPSKVDAT 288
QY 287 ----- 286
Db 289 DNRLVSGEGKAGKADGKGKDSKAIADTLAVTKESLDSTRENEELQSRMQDLSQLD 348
QY 287 -VEKPVKEKVA-----APEAKAEKPAVRPEPVPAAANTA 321
Db 349 KLQKLQLKDAQLAKLQGLGAGGQGAOPNAALPDASQPNAAQAQAPQGTPTAAAPT 408
QY 322 ASETAESAPOAAASAIIDPTDE--TGNVSEPVQVSAEETESGLFGSYTLLAG-- 378
Db 409 APAGEPAAPQPPVAPPAPAEKPPAPVAPVAPVQAAEQAPSPFL-----DELLNP 463
QY 379 -----GGAALIALLLLLLAQSKRARTTESVP-----EPEPDLDDAADGI-EITP 424
Db 464 LMLAVIGSALLALLVLLMLSRNQAQKEKEBAQAFADTGEQEEDALDLGKDFDGLTL 523
QY 425 AEVETPATPEPAPKXNDVNTLALDGESEEL-----SAKQTFDVTDTPSNRIDL 475
Db 524 DEPE-PQVAAVAPQVEKTTAQTSDALGEADIIYAYGRFNQAAELLQNAIYDEP-QRTDLR 581
QY 476 FDSLAA-AQNGILSGALTODEETQKADADNATSTDSVTEPEFNFYNEVEIVDTPE 534
Db 582 LKLMEVYAEIMGREGFAQENELRIGGQ-PQVQLKSRV-----PAMVAVAAVA 631
QY 535 PESVAQTAENKPEVDTDFDNLPSNNHIGTEASAKP-ASPSGLIAGFLKASSPETILE 593
Db 632 GLAGAKLAQDELDSFSLD-DLSLDDSGH-----AAKPDAGQDLDDAFDLSLDDLGD 683
QY 594 KTVASVQVPE-ELHDFLKVYETDAETAETPTPDNAAADDLSALLQPAEAPSVEENITE 652
Db 684 DVQADLKSDSGALDGLTLDLSDLLAASTPADKP-----VDDLDLDFGLDAE--- 728
QY 653 TVAETPDFNATADDLSALLQSEVPAVEENAAEIVADDLSALLQ---PAEAPAVEENYTE 709
Db 729 -LAETPS-OPKHDDLDGDFSLDADP--EDKUSD---DDFLSLNDEVEFAAAAFNEFTLD 781
QY 710 TVAETSDFTHTAAD--DLSALLQPAEVPVAVENKIVAIEIPDFNATADDLSALLQSEVP 767
Db 782 TEAAEPEALSUPDDFDLSLADEPTPEAPAEKGEDSFAQLDEVSQQLDELAS----- 833

```

```

QY 768 AVENAAEITLETDSNT---SEADALPDFLKOGEEETVDSIYLSEENIPNNADTSFPPS 824
Db 834 -----NLDEPKSATSPSADDAVASALDGD-----ADDDDFD 866
QY 825 ESVGSDAPSEAKYDLAEYMLIGDRDAAAETVQKLLLEAEAGDVLKRAQA 873
Db 867 LSGADERAT--KLDLARAYIDMGSEGAIDILDEVL--AEGNDSQQAEE 911

RESULT 7
Q9KTA5
ID Q9KTA5 PRELIMINARY; PRT; 1621 AA.
AC Q9KTA5;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Hypothetical protein VC0998.
GN VC0998.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
DR EMBL; AE004181; AAF94159.1; -.
DR PIR; A82255; A82255.
DR TIGR; VC0998; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1621 AA; 178073 MW; 2EF38BAC27A2485D CRC64;

Query Match 8.0%; Score 353.5; DB 16; Length 1621;
Best Local Similarity 23.1%; Pred. No. 1e-08;
Matches 226; Conservative 129; Mismatches 316; Indels 307; Gaps 50;

QY 137 REYTAILDPVGSPTKMSALSGKTHKTAPTAESQENQNAKALRKTOKDSANAARKPA 196
Db 24 RFFQRLILFVAVMVVVTQTSFVSAESIRLVGPDGVQVPTPQ---YSENIVRNSANN--EPG 78
QY 197 -YNGKTHTVRKGETVKQIAAARIP-KHLTLEQVADALLKANP---NVSAGRLRAGSVL 250
Db 79 RFFGPTSA---NQTLWSIASQLRPSSSVTQTLAIYQLNPQAPENQNIH-TLIFGSTL 134
QY 251 HIPNLNRIKAEQKPTQAKPKAETASMPSESKQATVEKPKPE-AKVAAP--EAKAEK 307
Db 135 RVPSLAQISRNSTQDAVNIMASHQAKINQTPD---TPVPVPAPRPAPVATPKVAVQAT 191
QY 308 PAVRPEVPAANTAASETAESAAPQAAASADTPT----- 343
Db 192 P---POVTPT-----TAPOEKAPTLEKTPAKPSQSTDAEYMALEEKNNHTLRML 237
QY 344 -----DETG--NAVSEPEVOVSABE-----ETESGLFGSYTLLLAGGG----- 380
Db 238 SQVQSEVSTLKEGLDENIRSEVERLLEERKAEASRLAPSLDNLNLSNGWLVALLA 297
QY 381 -----AALIALLLLRQAQSKARRTESVPPEEPDLDAADDGIEITFAEVETPATPE 434
Db 298 LIPGLLIATVLLNRRSSAQENPTNNITSEMP-----TAAPVTLG 341
QY 435 PAPKNDVNTLALDG-----ESEELSAKQTF-----DV-----ETDTPSN----- 470

```

```

Db 342 PEQTEIDGDDLLDDDLFSTTDDKEENDAEKAFSDEDDVPADLNETDLDFNLQGDQSDDL 401
QY 471 -----RIDLPDLSLAAQNGILSGALQTDE-----ETQKRA-----DADWNAI 508
Db 402 FVGIDDDGDLDTFEDALNESANGI---SVNADDKALGLEEMERALNDVSEPTDNDLNSF 458
QY 509 ESTSVVPEPTFNYPNVEIVIDTPEPE-----SVAQTAENKP----- 546
Db 459 DLADENQWSE-----DDIEALLSGDERNELLSGKVQSQSLDDLLASLDELDADEPAIQD 513
QY 547 -ETVDTDFSNLPSNNHIGTEETASAKPASPSGLAG-----FLKA 585
Db 514 TETLDTLLNDELASL-----SEEDDDEFDLSGAGVAGDQDLDDLFPASIEEQADLEQLRAKA 569
QY 586 SSPETILEKTVAEVQTP-----EELHDFLKVYETDVAETA-----PETPDFNAAA 631
Db 570 IDETALLDEILAEQDAPLSEESTELLDELLDDFKPENDEPDAQTAADLLOPEEPILDEE 629
QY 632 DDLSALIQ-----PAEAPS---VEENITETV-----AETPDFENATADDLSA 669
Db 630 DSTQLLNEVLGEVPPEELASGLEIDQNSTELLDELLDDLDDESIEATEFSVAPEKLSV 689
QY 670 -----LLQPSVEPVAEENAAEIVADD-----LSALLQPAEA-----PAVENV 707
Db 690 EDGTELFDELLEIEQHPESASLPELATEDEFNSDTFIDDLNLSGAPAKDPLLEPVLDENE 749
QY 708 TETVAETSDPHT-----AADLSALLQPAEVA-----VEENTKTV---AE 746
Db 750 APAQADDFDNFETEGGLEDDL---QPSALPANFETGPQDEDDWVDFDDSSPTLEGNAE 806
QY 747 IPDFNATADDLSALLQPSVEPVAEENAAEITLE---TPDSNTSEA---DAL-PDFLKDGE 800
Db 807 L-ELSSAEDDL-----PEQTATNETADELLADLAAQPSQNTVTSDDALLPDGLSQSV 860
QY 801 ETVDWS-IYLSEEN-IPNNAD---TSFSPSVSGSDAPSEAKYDLAEYMLIGDRDAAAE 854
Db 861 EPLTLDLELPEENDEPQLAEVTPSSAFDEQVETEIRPESEPLAAEASNDESILTALNE 920
QY 855 TVQKLLLEAEAGDVLKRAQ 872
Db 921 L--DLPEYTEEDVLADVQ 936

RESULT 8
Q8DB34
ID Q8DB34 PRELIMINARY; PRT; 1951 AA.
AC Q8DB34;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE AAA ATPase.
GN VV11991.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016803; AA010389.1; -.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
KW Complete proteome.
SQ SEQUENCE 1951 AA; 212759 MW; 73DBF9E1EED8DEE7 CRC64;

Query Match 7.9%; Score 347; DB 16; Length 1951;
Best Local Similarity 22.6%; Pred. No. 2.7e-08;
Matches 185; Conservative 137; Mismatches 305; Indels 192; Gaps 36;

```


SQ SEQUENCE 2768 AA; 294032 MW; CA929A21774E4684 CRC64;

Query Match 7.3%; Score 319.5; DB 5; Length 2768;
 Best Local Similarity 20.2%; Pred. No. 9.3e-07;
 Matches 198; Conservative 139; Mismatches 323; Indels 321; Gaps 40;

QY 101 TAKVHKLGDKAVIAVSQAVRDPVLVPR1-----GAGAQRVRE-YTA 141
 DB 1141 TAKPDNKIDVSEHISTEIEIPKDVIMPTGITEQLSHVHPDEIQTPTSVPAQDESITTA 1200
 QY 142 ILD--PVGYSPKTSALSADGKTHRTAPTASQEN-----QNAKALRTDKDSANA 191
 DB 1201 KVDKKPIDESAEDKKPI--GESEDSKPIDSEEDKKPVEASAEDEKPEVEDSEKPELP 1258
 QY 192 AVKPAYNGKTHVRKEIVKQIAAAI-----RPKHLTLQVADALLK-----ANPNVSA 240
 DB 1259 TVIPA--SEIBKESKPEDEKTEADFAAPTQOPEATTQAQADTAKEVDDKLATTSAPV 1316
 QY 241 HGLRAGSVLHPLNLR1KAQPKQTAQKPAETASMPSEPSQATVKEPKPEAKVAA 300
 DB 1317 SGE-----DELKPADEKKRTETAIP-----DAEIPASTDEPESSTEL 1354
 QY 301 PEAKAEK-----PAVRPEPVPAANTAAAS-ETAABE-----APQEAASAIITP 342
 DB 1355 PTVDLDDKPEEDSTKGTAPESDKVPEVPTASTENEIEESDKFTTVAPPKISADETEP 1414
 QY 343 TDETC--NAVSEPVQO-----VSAAEETESGLFGGSY 372
 DB 1415 TAEDLVATPEPISESEVSTKPAVQGPPLPTLAPQPEKKPVDAITSTADI----- 1469
 QY 373 TILLAGGGAALIALLLLR1LAQSKRAARTEESVPEEPDLDDADDG-----IETFAEV 427
 DB 1470 -----STEPSAEVEKASGETSDNEIDAGASSTVPVPSADED 1508
 QY 428 ETPATPEPAPKNDVNDTIA-LDGESEBELSAQTTDV-ETDTPSNRIDLDFSLAAQNG 485
 DB 1509 KTPSTEKTVEADDKFTTVAPLAGDEESNLPKLPQDIFEEAPV-----A 1553
 QY 486 ILSGALTODEETQKRAADNNAL1ESTDSVYBETFPNVN----- 524
 DB 1554 VTTAAPSCKDGEQKPVVEEKEPIEDGQKPIEDTSTPTSSSENEIPESDRATTIAPSKEE 1613
 QY 525 -----PVEIVIDTPEPSVAQTAENKPTVDVTDPSDNLPSNNHIGTTEE-TAS 570
 DB 1614 PSEPSTGAPTCKDEPAEPSTDAPESDSKETPESEVPTTVAPAGEKIPTSSITPDEPTAT 1673
 QY 571 AKPASPSGLAGFLKASSPTILEKTVAEVQTPEELHDFLKVYETDAVATAPETPDFNAA 630
 DB 1674 SAP-----VAKPDEDEK-----ETSTEI-----PTDAPASS-----BEDENS 1707
 QY 631 ADDLSALLQPAEAPSVEENITETVAETPDFNATADDLSALLQPSVPAVEENAAETVADD 690
 DB 1708 TQOI-----PSEVP-----EKKPTPAQTPEGDIVGNATAPTTSDEVPVQRLPEVLAIE- 1758
 QY 691 LSALLQPAEAPAVEENVTVTAFTSD-----PHTAADLDSALLQP----- 730
 DB 1759 ---IPQSTETG1KQDQDTEAAPSIDRKEPVYVEIDEETVAP1SEKDEKPTEEKPVYE 1815
 QY 731 -----AEVPAVEENVTKV-----AE1PDFNATADDLSALLQPSV-PAVEENAAETITET 780
 DB 1816 OKPTGEPSSEEEKEKPIEQDVSTEGP-VSTEASEAGSTESSEVEKPESTEGEVAEKPEDK 1874
 QY 781 PDSNTSEA--DALPDF-----LKDGEEETVDMISIVLSEENIPNNAOTSPPSE----- 825
 DB 1875 QPSSTAQAQVETIPEISTELPAQDGKPT-----SEAPVDSDETSAPSDEKIPSVSG 1927
 QY 826 -----SVGSDAPSEAKYDLAEM-----YLETIGDRDAAAEATVQKLLLEA 863
 DB 1928 EEVEGPEVTTASPOAAEEDLKTPAESEPSTDKVPETEQKPEDETKADETPESVTVQS 1987
 QY 864 E-----GDVLKRAQA 873
 DB 1988 DVATSTAPVAGGDIEKDEQA 2008

RESULT 13

Q07290 PRELIMINARY; PRT; 1822 AA.

AC Q07290;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Epf* protein.
 GN Epf*.
 OS Streptococcus suis.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1307;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1890;
 RX MEDLINE=93328288; PubMed=9335363;
 RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
 RT "Repeats in an extracellular protein of weakly pathogenic strains of
 RT Streptococcus suis type 2 are absent in pathogenic strains.";
 RL Infect. Immun. 61:3318-3326(1993).
 DR EMBL; X71880; CAA50714.1;..
 DR PIR; S33441; S33441.
 DR GO; GO:0016020; C:membrane; IEA.
 DR InterPro; IPR005877; Gpob_Ysirk.
 DR Pfam; PF04650; Ysirk signal; 1.
 DR TIGRFAMs; TIGR01168; Ysirk signal; 1.
 DR SEQUENCE 1822 AA; 192632 MW; 3838960C77641D7D CRC64;

Query Match 7.2%; Score 316; DB 2; Length 1822;
 Best Local Similarity 19.9%; Pred. No. 7.8e-07;
 Matches 204; Conservative 165; Mismatches 363; Indels 294; Gaps 43;

QY 37 NNRQIKLTAASV-----AVAAAPQAHAGLGLNIQSNLD--EPFSGSITVTGEAKAL 87
 DB 682 DNERLKLGPDSFTVNSDGTVSVDYSA-----GGVNDGATDIKNAATNLADTRQAKAE 737
 QY 88 LGGGSVTVSEKGLTAK-----VHKLGDKAVIAVSQAVRDPVLVFR1GAGAQRVREYTAIL 143
 DB 738 I-DTKLAHEKKAIEAKRDEAFSKIDDISLRAEQQAADKAAVAA--AAGDALKE-----L 789
 QY 144 DPVGVSPK-----TKGALSDGKTHRTAPTAESEQENQAKALRTDKDSANAAY--- 193
 DB 790 DNKATEAKEKIDKATTASEINDAKTNGEI--NLDSEAEVGEKAIINQAKELAKARVENK 847
 QY 194 -----KPAY-----NGKTHVRK--GETVQ1AAAIRPKHLT----- 223
 DB 848 APEALEKNNPNLLEBKAYPDDIKESKEVAVKINNAENTAETITAAIDEAETAYNED 907
 QY 224 -----LEQVAD---ALLKANPNVSAHRLRAGSVLH1PNLNRIKAEQPKQTA 268
 DB 908 V1NAAQLDALNKLKXDEBETKAAIDANPNLTPEEKAKA-----IAKVEELVNAESDILS 962
 QY 269 KPAETASMPSPSQATVKEPKPEAKVAAPEAK---AEKPAVRPEVPAPANTAASET 325
 DB 963 KPTPTVQAVEDKA-----DKDLAKVELQAAADGAKKGEIAPENPTPEKDVAKKAVEDA 1017
 QY 326 AAEAPQEAASAIID---TPTDETGNVSEPVQVSAEE--ETESGLFGSVTL1LAGGG 380
 DB 1018 V-----KVATDAIDKASTPT-EVDTATSDGVKADAEAEFKATQKD----- 1056
 QY 381 AALIALLLLR1LAQSKRAARTEES-----VPEEPDLDDADDGIEITFAVETP 430
 DB 1057 -----AKNKIAKEAESAKKAIIDNPNLTPEKESAKNAVEEAARVATAIDKA 1104
 QY 431 ATPEPAPKNDVNDTLALDGESEBELSAKQTFDVEITP-----SNRIDLDFSLAAQ 483
 DB 1105 STPD-----AVQVEEDKGVAAILN1ITAKADAKGVIAKLADEIKKLEKQAE 1153
 QY 484 NGTILSGALTODEET-----OKRADADNNA1ESTDSVYBETFPNPFVEI----- 528

```

Db 1154 KADASTWNEEKAIAKALQDVVDKGALEDAARVATNEIIBATTTEKAKAEALAGEK 1213
QY 529 -----VIDTPESVAOTAENKPTVDTFDSNLPNNHIGTEETASAKPASPS 577
Db 1214 SLTDTGKEARDAVELAKDELAEIIRTEEBEATKIVEKLAEDTRKAIENPNLSDEKQ 1273
QY 578 GLAGFLKASSPFTI-----LEKTVAEVQTPPEELHDFLKVYETDAVAETAPETPDFNAAA 631
Db 1274 AEIKKLTDAVAKTLATIRDNADKRTQEAERKAQALADLEKAKETOKIAD-----KAAI 1325
QY 632 DLSALLOPAEAPSVEEN-----LITEVAETPDFN-----ATAD 665
Db 1326 DRITILVKCELEATKQDKNKIADKAAAKEAIAENPNLTDAEKKTTFTDAVDAEVAKAN 1385
QY 666 D-LSALLQPSVEPAVENNAEIVADD-LSALLQPA-----EAPAVEE-----NTE-- 709
Db 1386 DAISAATSPADVQKEEDAGVAAAEVDLDRADKQDKNIAKADAAKEAIGSPNLTDAE 1445
QY 710 -----TVAETSDFTHTADDLSALLQPAEVPVAVENVTKTVAEIPDFNATADDLSALLQPS 765
Db 1446 KKTFTDAVDAEAVAKANDAIASAATSPADVQKEEDAGVAAIAE-----DVLDAAKQDAK 1497
QY 766 VPAVENNAEITLETDPDNTSEADALPDFLKDGEEETVDWSIYLSEENIPNADTSPSE 825
Db 1498 NTKAKES-----DAKSAIDANFN-LTDAEKESAKKAV-----DADAKAATD 1538
QY 826 SV-GSDAPSEAKY-----DLAEMYLEIGDRDAAAEATVOKL-----LEAEAG 865
Db 1539 AIDASTPVEAQSAEDKGVGSIADQVLDRAKQDKNIAKEVAEADANFNLSDAEK 1598
QY 866 DVLKRA 871
Db 1599 EASKKA 1604

RESULT 14
Q9UOP0 PRELIMINARY; PRT; 1786 AA.
AC Q9UOP0
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
RL protection in chimpanzees";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ007010; CAB65343.1; --
DR GO; GO:0003723; F:RNA binding; IEA.
DR InterPro; IPR001313; Pumilio/Puf.
KW Signal.
FT SIGNAL 1 63 POTENTIAL.
FT CHAIN 64 1786 POTENTIAL.
SQ SEQUENCE 1786 AA; 200101 MW; 5DF536D7B5B1BD98 CRC64;

Query Match
Best Local Similarity 20.2%; Pred. No. 8.1e-07;
Matches 167; Conservative 147; Mismatches 331; Indels 183; Gaps 26;

QY 64 NIOSNLDPEFFSGSIITVGEAKALLGGGVTVSEKGLTAKVHKLQKAV---IAYSSEQA 120
Db 242 NVEENVEENDGVSASSVEES-----IASSVDESIDSSIEENVAFTVEEI 286
QY 121 VRDPVLVFRIGAGAVREYATLDPVGYSPKTSALSODGKTHRKTPATAES---QENQAK 178
Db 287 VAPSVV-----ESVAPSVSESVSEENVEES 310

```

```

QY 179 ALRKTCKDSANAAVKPAYNGKTHTVRKGE-----TVKQIAAAAIRPKHLTLEQ-VADALL 232
Db 311 VAENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP-----TVEEIVAPSVV 364
QY 233 KA-NPNVSAHGRILRAGSVLHIPNLNRIKAPQPKQTAQKPAETASMPSEPSKQATVEKPV 291
Db 365 ESVAPSVSEVE-----ENVEESVAENVEESVAENVEESVAENVEESVAENVEESV 415
QY 292 EKPEAKVAPE-----AKAEKPAVRPEPVAANTAASETAESAPQAAAAIDTPTDET 346
Db 416 AENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVSESVSEENVEESVAENVEESVAENV 475
QY 347 GNAVSPVEQVSAEETESGLFCGSGTYLLLAGGGAIALILLRLAQSRAARTTESVP 406
Db 476 EESVAENVEESVAENVEES-----VAENVEESVA 504
QY 407 EE-EPDLDDAADGCIITFAEVTPTATPPAPKPNVDNLTALDGESEELSAKQTPDET 465
Db 505 ENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVSESVSEENVEE-----NVEE 557
QY 466 DTPSNRIDLDFSLAAQNGIISGALTQDEETOKRADAMNAIESTDSVYETPTFNPYP 525
Db 558 SVAEN-----VEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP 601
QY 526 -VEIVIDTPESVAQTAENK-PEYVDTDFSDNLPNNHIGTEETASAKPASGLAGFL 583
Db 602 TVEEIVAPSVVSVAPSVSESVSEENVEESVAENVEESVAENVEESV----- 647
QY 584 KASSPETILEKTVAEQTPPEELHDFLKVYETDAVAETAPETPDFNAAADDLSALLOPAEA 643
Db 648 -AENVEEIVAPTVEEIVAP-TVEEIVAPSVVSVAPSVSESVSEENVEESVAENVEESVA 704
QY 644 PSVEENITETVAETPDFNATADDLSALLQPSVPAVEENAAEIVADDLSALLQPAEAPAV 703
Db 705 ENVEESVAENVEE-----SVAENVEEIVAPTVEEIVAPTVEEIVAPSVVSVAPSVSESV 759
QY 704 EENVTETVAETSDFTHTADDLSALLQPAEVPVAVENVTKTVAEIPDFNATADDLSALLQ 763
Db 760 EENVEESVAENVE-----ESVAENVEESVAENVEESVAFTVEEI-----VAPSVSE 805
QY 764 SEVPVAENNAEITLETDPDNTSEADALPDFLKDGEEETVDWSIYLSEENIPNADTSP 823
Db 806 SVAPSVSEESVA-----ENVATNLS-D-NLLSNLGGTETEIKDSILNEIEEVENVTITI- 859
QY 824 SESVGSADPSEAKYDIAEMYLEIGDRDAAAEATVOKLLEAEAGDVLKRA 871
Db 860 LENV-EETTAESVTTFSNILEBIQENTITNDTIEEKLLEHENVLSAA 906

RESULT 15
Q8XTC4 PRELIMINARY; PRT; 673 AA.
AC Q8XTC4
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Probable transmembrane protein.
GN RSP0189 OR RS04701.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chardier M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,

```

Search completed: September 2, 2004, 19:20:43
Job time : 156 secs